

Izvirni znanstveni članek ■

## Interaktivne spletne aplikacije v biomedicini: primer dictyExpress-a

## Interactive Web Applications in Biomedicine: the dictyExpress Case Study

**Blaž Zupan, Gregor Rot, Tomaž Curk**

**Izveček.** Dinamične interaktivne spletne strani so lahko zelo primerne za razvoj aplikacij za podatkovno analitiko. Grafični vmesniki in načini interakcije, ki temeljijo na že ustaljenih pristopih iz namiznih aplikacij, lahko uporabniku spletnih analitičnih sistemov pomagajo pri enostavnem dostopu in brskanju po spletnih bazah ter mu olajšajo iskanje zanimivih vzorcev. V članku opišemo primer razvoja tovrstnega sistema, imenovanega dictyExpress, ki je eden od znanilcev novega vala interaktivnih spletnih aplikacij na področju biomedicine.

**Abstract.** Dynamic interactive web applications can nicely support data analytics. Their foundations are graphical interfaces and user interactions that resemble those from desktop applications. Due to simplicity for the users, such web applications may provide a substantial support for data retrieval and navigation through public databases, and help in mining interesting data patterns. An example of such application, called dictyExpress, is described in the paper. Due to its rich graphical interface and many modes for user interaction we see it as a precursor of a new wave of biomedical web applications.

■ **Infor Med Slov:** 2010; 15(1): 10-17

---

Institucija avtorjev: Fakulteta za računalništvo in informatiko, Univerza v Ljubljani.

Kontaktna oseba: Blaž Zupan, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Tržaška 25, 1000 Ljubljana. e-pošta: blaz.zupan@fri.uni-lj.si.

Prejeto / Received: 23.06.2010  
Sprejeto / Accepted: 29.06.2010

## Uvod

Najbrž so prav internetne aplikacije tiste, ki so še posebej zanimive za raziskovalce s področja biomedicine. Te navadno ponujajo dostop do posamezne zanimive zbirke podatkov, morda na privlačen način združujejo in prikazujejo podatke iz različnih virov, uporabniku nudijo enostavne možnosti dostopa do informacij, ter, za razliko od standardnih namiznih aplikacij, ne potrebujejo nikakršne namestitve ali posebnega vzdrževanja. Zanimive so tudi za razvijalce, saj morajo ti vzdrževati eno samo (strežniško) verzijo aplikacije in jim ni potrebno skrbeti za celo vrsto mehanizmov, ki pri uporabnikih poskrbijo za namestitve in uporabo zadnje verzije programa. Tudi testiranje tovrstnih aplikacij je poenostavljeno, saj razvijalcem ni potrebno predvideti številnih možnih nastavitvev na osebnih računalnikih in uporabniških delovnih okolij.

Težave pa nastopijo pri podpori interaktivnosti. Na to smo pri namiznih aplikacijah navajeni. Na primer: dvoklik na ikono odpre ustreznih dokument. Klik na črto v grafu izbere točko (podatke), ki ta del grafa določajo. Izbor nekega objekta na sliki omogoča nastavitve parametrov njegovega prikaza. Tipične spletne aplikacije so veliko bolj statične. Izbiramo med nekaj vnaprej pripravljenih možnostmi (spletne povezave) ali pa odgovorimo na vprašanja v predpripravljenem spletnem obrazcu ter potem zahtevamo naslednjo stran ali osvežitev prikaza s klikom na gumb "Naprej". Tovrstne spletne strani so statične, čeprav je njihova vsebina lahko odvisna od vnesenih podatkov. Statičnost izvira iz načina dela aplikacij na spletu: spletne strani izrisuje brskalnik, ki opis strani pridobi od strežnika. "Lene", statične spletne aplikacije za vsako spremembo strani kličejo ustreznih program na strežniku, ki na podlagi uporabnikovega vnosa vrne novo spletno stran. To brskalnik izriše, uporabnika spet vpraša za izbiro nove strani ali mu ponudi obrazec za vpis podatkov. Po uporabnikovi akciji brskalnik spet povpraša strežnik po novi strani in tako naprej. Delo s tovrstnimi aplikacijami je popolnoma drugačno kot z namiznimi aplikacijami,

interaktivnost pa pravzaprav ni podprta (če sem ne prištevamo izpolnjevanja obrazcev in klikanja na pripadajoče gumbе).

Statičen način dela je danes prisoten v večini najbolj znanih spletnih aplikacij v biomedicini, čeprav so te razvijalci v preteklih letih zelo osvežili ter jim nadelali nov, privlačnejši izgled. Sem sodijo strani PubMed ([www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed)), vsa ostala paleta NCBI-jevih spletnih aplikacij ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), aplikacije Evropskega Inštituta za Bioinformatiko, kot je sicer izjemno koristni BioMart ([www.biomart.org](http://www.biomart.org)), ter vmesnik najnovejših bioinformatičnih prizadevanj kot je Galaxy ([main.g2.bx.psu.edu](http://main.g2.bx.psu.edu)). Omenjene namreč slonijo na tehnologijah, ki izvirajo iz protokola CGI (Common Gateway Interface), kjer se celotna vsebina strani generira na strežniku in mora po njej po vsakem vpisu podatkov povpraševati odjemalec – brskalnik. Neglede, ali razvijalci uporabljajo PHP, Java Server Scripts, strežniške skripte v Perlu ali Pythonu, je rezultat enako vizualno in uporabniško okoren.

Seveda se bo to trenutno stanje kaj kmalu (kot ostale zadeve v računalništvu in na področju informacijske tehnologije) zelo spremenilo. Prihajajo namreč interaktivne aplikacije, ki uporabljajo tehnologije dinamičnih spletnih strani – tako kot to počno npr. aplikacije Googleove spletne pisarne. Dinamične spletne strani temeljijo na tehnologijah Flash, Java Script, Ajax in HTML 5. Vsebinsko strani določa program, ki se izvaja znotraj brskalnika, torej pri odjemalcu. Če je potrebno, ta za podatke ali le dele strani povpraša strežnik. Vse to se zgodi ne da bi uporabnik opazil, še manj pa da bi mu bilo potrebno klikniti na temu namenjen gumb. Dinamične spletne strani posnemajo grafične uporabniške vmesnike, ki smo jih sicer navajeni na namiznih aplikacijah, a ob tem zadržijo vse prednosti, ki smo jih uvodoma omenili.

Dinamične spletne strani so izjemno pomembne za razvoj aplikacij v biomedicini, ki največkrat slonijo na obdelavi, prikazu in uporabi podatkov, kar pa je bistveno olajšano, če je podprto s sodobnim grafičnim vmesnikom, ki uporabniku omogoča povpraševanje prav po vsemu kar je prikazano na

zaslonu. V primerjavi s statičnimi spletnimi stranmi, kjer je tip interakcije omejen na klikanje po vnaprej predvidenih povezavah ali pa izpolnjevanju spletnih obrazcev, ponujajo dinamične spletne strani veliko bogatejši nabor funkcij, ki jih uporabnik lahko proži takrat, ko jih potrebuje in ne takrat, ko mu jih sistem pač ponudi.

Interaktivnih spletnih strani na področju biomedicine je danes še izjemno malo. Razlogov je več. Tehnologija in standardi so šele v nastajanju in jih različni brskalniki različno dobro podpirajo. Tudi razvojna orodja šele sledijo razvijajočim se standardom in za dinamične spletne strani niso še tako udobna kot sorodna orodja namenjena razvoju klasičnih namiznih aplikacij. Nenazadnje, razvojnega kadra izrazito primanjkuje oziroma se teh tehnologij šele uči.

Ne primanjkuje pa problemov, ki bi jih z izdelavo dinamičnih, interaktivnih spletnih strani lahko rešili. V nadaljevanju članka opišemo razvoj slovenske spletne rešitve, za katero mislimo, da je zgledni primer prihajajočih spletnih aplikacij. Aplikacija, imenovana dictyExpress ([www.ailab.si/dictyexpress](http://www.ailab.si/dictyexpress)),<sup>1</sup> je namenjena molekularnim biologom – raziskovalcem socialnih ameb iz družine *Dictyostelium*, ki jim na zanimiv, interaktiven način ponuja podatke o izraznih profilov genov divjega tipa in izbranih enojnih in dvojnih mutantov.

## Pred dictyExpress-om

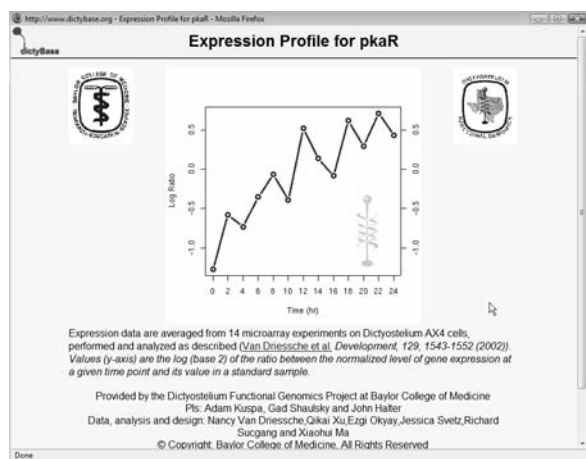
Amebe *Dictyostelium*, med katerimi je najbolj raziskana vrsta *Dictyostelium discoideum*, so prav posebna živa bitja.<sup>2</sup> Živijo tik pod površino prsti in se prehranjujejo z bakterijami. Prav dosti se ne menijo druga za drugo, vse dotlej, dokler ne zmanjka hrane. Takrat najbolj lačne prično oddajati kemične signale, ki jih sprejmejo druge amebe. Te se usmerijo v smer največjega gradienta signala, torej proti amebi, ki je pričela signal oddajati, ter se v nekakšnem spiralastem plesu združijo v črvu podobno tvorbo. Več 100.000 ali celo milijon jih tako skupaj potuje proti svetlobi in

toploti (površju), potem pa pride do nenadne diferenciacije: "črv" se pretvori v bazalni disk (oprijemališče), dolg pecelj in sporo. Približno sedmina ameb oleseni ter se tako žrtvuje za amebe znotraj spore, ki jih voda ali kakšen drugi mehanski transport prenese na drugo rastišče, kjer je hrane dovolj. Spora tam razpade, cikel rasti in razvoja se ponovi.

Celotni razvojni cikel amebe traja en dan in je zelo robusten. V procese, ki ga spremljajo, je vključena večina od 12.500 genov tega organizma.<sup>3</sup> Nekateri se izrazijo prej, na začetku razvoja, nekateri na sredini, nekateri ob zadnjih fazah. *Dictyostelium* je prav zaradi tega hitrega in enostavno ponovljivega razvojnega cikla postal izjemno privlačen modelni organizem, s katerim danes raziskovalci proučujejo vlogo posameznih genov. Pri tem lahko s tehnikami DNA mikromrež spremljajo genske izraze več tisoč genov naenkrat. Te merijo v več različnih časovnih točkah, na primer vsaki dve uri razvojnega cika, ter na ta način pridobijo genske časovne profile.

Z merjenjem tovrstnih časovnih izraznih profilov sta se na svetu ukvarjali dve instituciji iz ZDA, Baylor College of Medicine v Houstonu in University of California v San Diegu. Da bi preostalim raziskovalcem omogočili dostop do pridobljenih eksperimentalnih rezultatov, so profile objavili v grafični obliki na portalu organizma *Dictyostelium discoideum* ([www.dictybase.org](http://www.dictybase.org)).<sup>4</sup> Do izraznih profilov za posamezni gen je bilo moč priti na domači strani tega gena s klikom na ustrezno povezavo. Ta je odprla okno, katerega primer za gen *pkaR* prikazuje slika 1. Numerične podatke, ki so osnova takemu grafu, je bilo sicer moč dobiti v dodatnem gradivu ob objavljenih člankih,<sup>3,5</sup> ali pa v standardnih odlagališčih z izraznimi podatki (npr. Gene Expression Omnibus). Vsa dodatna analiza je bila prepuščena iznajdljivim programerjem, ki so take podatke morali pridobiti iz spletnih baz, jih v posebnem programu urediti, obdelati in na ustrezen način prikazati. Biologi, tudi ti, ki so sicer zbrali tovrstne podatke, so bili na ta način odrezani od njih, in bili za potrebe še tako enostavne

podatkovne analitike vsakokrat ter popolnoma odvisni od kolegov informatikov in programerjev.



**Slika 1** Grafična informacija o genskem profilu izbranega gena, kot jo je pred uporabo dictyExpress-a ponujala spletna stran [www.dictybase.org](http://www.dictybase.org).

## Pričakovanja

Projekti, praviloma financirani iz javnih sredstev, ki zberejo tako veliko množico eksperimentalnih podatkov, so seveda primerno dragi. Financer praviloma zahteva, da so po končanem projektu podatki objavljeni javno, v spletno dostopnih zbirkah, a zaenkrat še ne vztraja pri objavah podatkov v okoljih, ki uporabnikom nudijo enostaven dostop in osnovno, če ne že napredno podatkovno analitiko. Želja raziskovalcev, ki te podatke pridobijo, seveda je, da bi bili ti čim bolj in čim širše uporabljani. Zgoraj opisan način dostopa do podatkov o izraznih profilih *Dictyosteliuma* tega seveda ni zagotavljal. Zato smo približno pred dvema letoma v skupini članov sedanjega Laboratorija za bioinformatiko na Fakulteti za računalništvo in informatiko v Ljubljani ter članov Laboratorija Gadija Shaulskega iz Baylor College of Medicine pričeli oblikovati aplikacijo, ki bi uporabnikom ponudila popolnoma drugačno izkušnjo in jim omogočila enostaven dostop do podatkov ter grafično podprto podatkovno analitiko. Slednja bi uporabnikom zagotavljala enostaven dostop do

različnih vizualizacij podatkov in možnost iskanja vzorcev oziroma poljubnih izraznih profilov.

Načrtovanje aplikacij je na področju informatike sicer odlično podprto in standardizirano, a ker gre za relativno enostavno aplikacijo, teh orodij nismo uporabili. Nasprotno, pričeto načrtovanje je bilo popolnoma neformalno, ob listu papirja, kamor smo skupaj risali pričakovan grafični vmesnik ter si ob skici izmišljevali morebitne funkcije, ki bi jih sistem lahko podprl. Pri tem smo tako razvijalci kot uporabniki morali odmisli, da gre za spletno aplikacijo, ki bo močno odvisna od podpore ustreznih programov na strežniku, in program načrtovati izključno s stališča uporabniškega vmesnika. V procesu razvoja dictyExpressa je bil tak način načrtovanja prav gotovo ključni dejavnik za uporabnost in popularnost aplikacije.

DictyExpress naj bi na enostaven način omogočal izbor eksperimenta (divji tip, mutirani sev ali ob določeni učinkovini opazovani organizem) ter genov, za katere bi zahtevali prikaz profilov. Podpiral naj bi:

- izbor vseh genov, katerih ime se prične z določenim znakovnim nizom,
- izbor vseh genov, ki imajo profil podoben določenemu ali izbranemu referenčnemu profilu,
- izbor vseh genov z določeno funkcijo, skladno z gensko ontologijo,<sup>6</sup>
- prikaz skupin genov v dendrogramu (uporaba hierarhičnega razvrščanja v skupine) ter možnost izbora veje dendrograma oziroma njej pripadajočih genov,
- prikaz mreže podobno izraženih genov ter v njej izbor referenčnega gena in njegovega profila,
- sočasni prikaz izraznega profila za izbrani (referenčni) gen v različnih eksperimentih,

- prikaz genske ontologije in z izbranimi geni obogatenih genskih skupin,

Vse vizualizacije v dictyExpressu naj bi omogočale interaktivnost v smislu izbora določene prikazane komponente oziroma pripadajočih genov oziroma eksperimentov. Dodatna želja uporabnikov je bila, da aplikacija ohranja svoje stanje, to je, da ob njeni naslednji uporabi brskalnik prikaže stanje ob zadnjem obisku strani. Stanje aplikacije z vsemi pripadajočimi nastavitvami naj bi bilo moč shraniti in ga kolegu poslati po elektronski pošti. Do aplikacije naj bi bilo moč dostopati preko povezave, ki vsebuje podatke o tem, kaj naj aplikacija prikaže. Slednje naj bi bilo uporabno za vključevanje aplikacije v referenčne portale, kot je npr. dictyBase.<sup>4</sup>

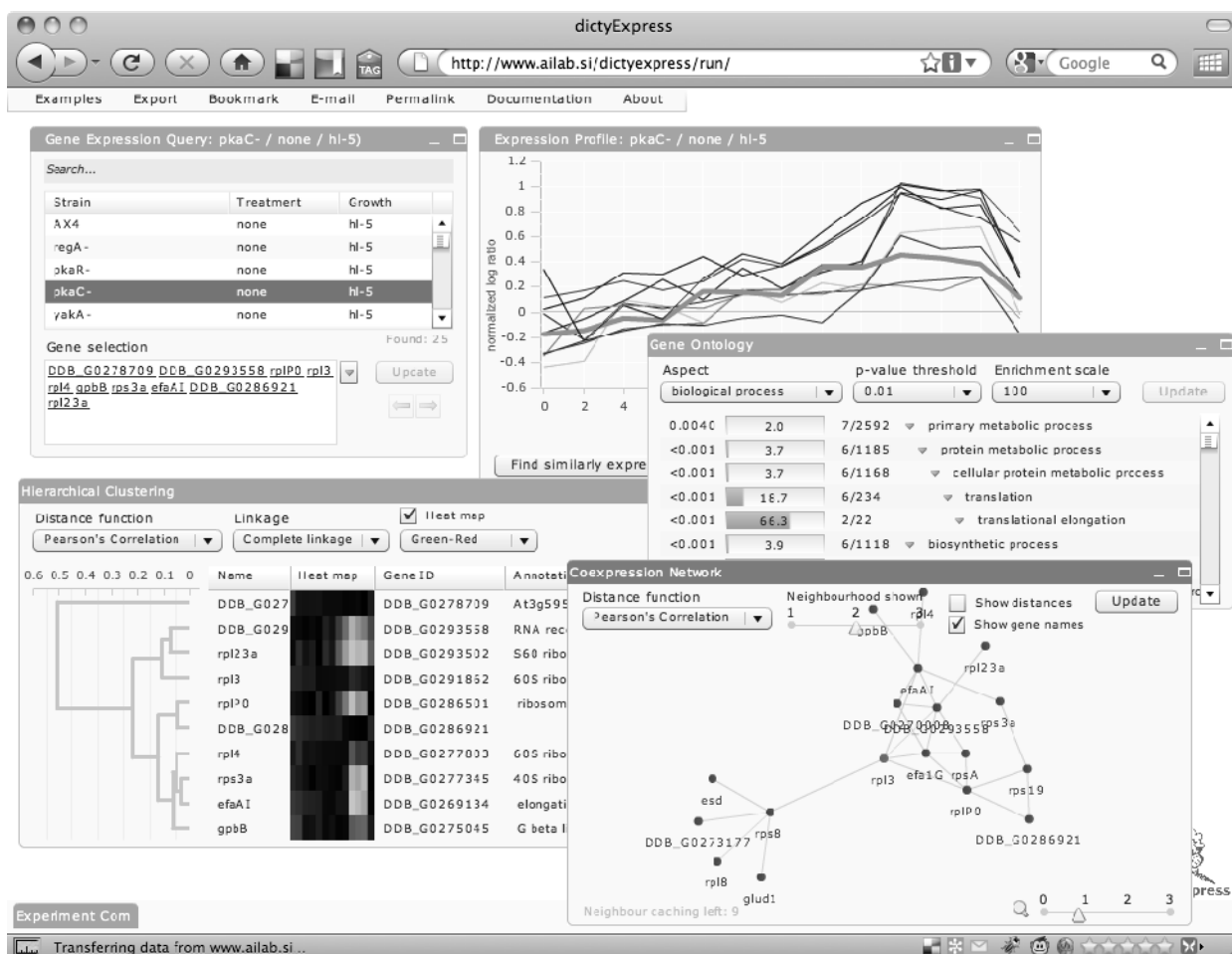
Aplikacija naj bi delovala odzivno. Uporabnik naj za posamezne zahteve ne bi čakal več kot sekundo, največ dve. V praksi je to pomenilo kombinacijo implementacije odjemalca, ki v glavnem podpira grafični interakcijski vmesnik, z učinkovito strežniško aplikacijo, ki podpira prav vse numerično zahtevne operacije podatkovne analitike. Seveda na način, kjer se uporabnik te komunikacije ne zaveda, kjer je potrebno optimizirati količino prenesenih podatkov, in kjer je nujna časovna optimizacija strežniške aplikacije,

tudi s predčasnimi izračuni za nekatere pričakovano pogoste operacije.

## Izpolnjene obljube

Prototip dictyExpressa je bil razvit relativno hitro, v nekaj mesecih, tudi zahvaljujoč odlični razvojni podpori, ki jo nudi Adobe Flex, ter delitvi programerskega dela na podatkovni, analitični in uporabniški del, ob vnaprej dogovorjenih standardnih za izmenjavo podatkov, za katerega smo uporabili osnovni protokol HTTP. Optimizacija celotne aplikacije, dodajanje nekaterih novih uporabniških funkcij, testiranje in razhroščevanje, izdelava spletnih strani z dokumentacijo, uvajalnimi video posnetki in primeri uporabe pa so zahtevali še dodatno leto dela.

Dokončana aplikacija, ob času pisanja tega prispevka dostopna v verziji 1.5, je prikazana na sliki 2. Glavne komponente aplikacije so: izbor eksperimentov in genov, prikaz izraznih profilov, hierarhično razvrščanje v skupine, obogateni deli genske ontologije in mreža podobno izraženih genov.

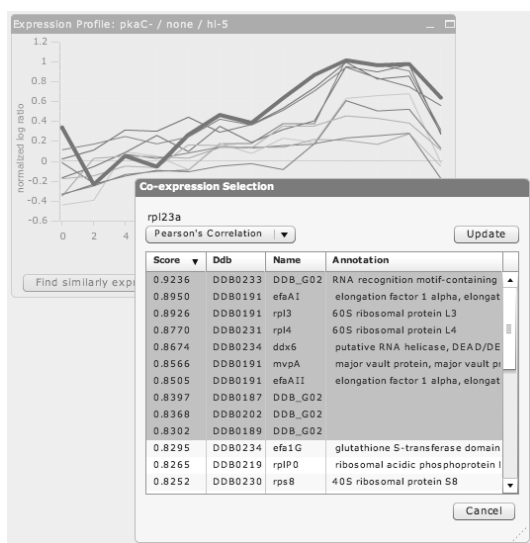


**Slika 2** Glavne komponente aplikacije dictyExpress. Sprememba v katerikoli komponenti, na primer, izbor genov v komponenti s hierarhičnim razvrščanjem v skupine, se hipoma prenese v ostale komponente, ki temu primerno osvežijo svoj izgled.

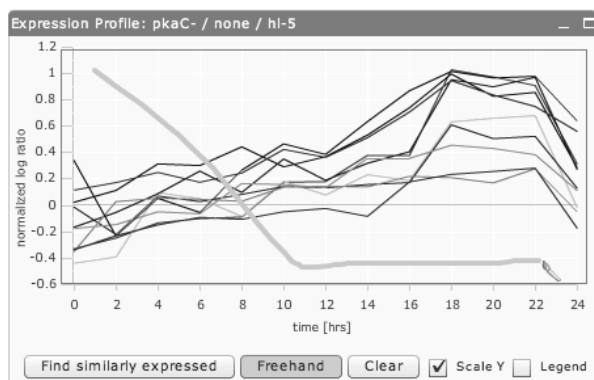
Na zaslonskih izrisih je seveda težko demonstrirati interaktivnost in dinamičnost aplikacije, a poskusimo vseeno. Slika 3 prikazuje izrazne profile, med katerimi smo izbrali enega (referenčnega) in zanj zahtevali izpis podobnih profilov, kjer podobnost izračunamo na podlagi korelacije. Uporabnik je izbral prvih deset genov. Potrditev izbora bo dictyExpressu naročilo osvežitev vseh ostalih oken z grafičnimi prikazi tako da bo uporabljena nova skupina enajstih genov, skupaj z referenčnim.

Referenčni profil lahko tudi prostorsko narišemo (slika 4). V dendrogramu lahko uporabnik izbere vejo s podobnimi profili ter zahteva prikaz informacij samo za izbrane gene (slika 5). Prikaz izraznih profilov za izbrani gen, a različne eksperimente (slika 6), omogoča izbor profila oziroma eksperimenta, po katerem se za dani nabor genov vse grafične predstavitve osvežijo in upoštevajo novi eksperiment.



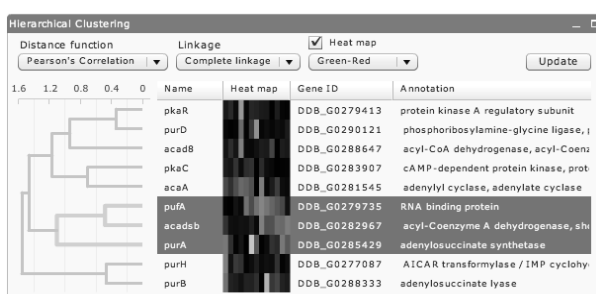


**Slika 3** Izbor referenčnega izraznega profila (odebeljen, v ozadju) in pregled najbolj podobnih profilov. Uporabnik je izbral prvih deset najbolj podobnih genov.

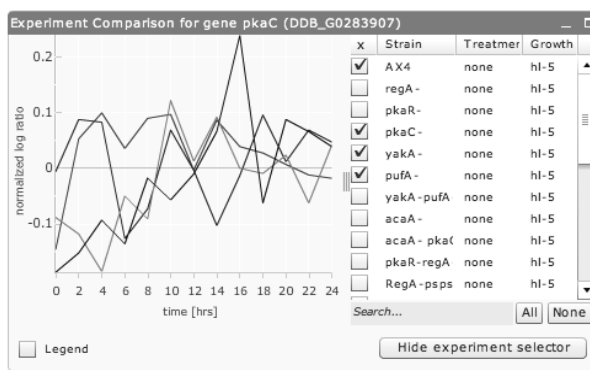


**Slika 4** Ročni izris referenčnega izraznega profila.

Na ta in podobne načine smo zadostili prav vsem željam skupine uporabnikov, ki je sodelovala pri razvoju. Na posnetkih zaslonov seveda nismo prikazali vseh funkcij. Na primer, uporabnik lahko shrani izrazne profile za dani izbor v lokalno datoteko, ki jo nato uvozi v namizni program za delo z razpredelnicami. Implementirana je tudi opcija »Undo/Redo«, s katero se uporabnik sprehaja nazaj in naprej po prejšnjih izborih genov in eksperimentov. Trenutni izbor in zgodovina izborov ostanejo zapisani tudi po prenehanju uporabe spletne aplikacije in so na voljo uporabniku ob ponovnem obisku.



**Slika 5** V denrogramu – prikazu hierarhičnega razvrščanja genov v skupine glede na podobnost izraznih profilov – je moč izbrati skupino podobno izraženih genov.



**Slika 6** Prikaz izraznih profilov enega samega gena v različnih izbranih eksperimentih.

Aplikacijo dictyExpress gradi baza podatkov (te v tem članku nismo posebej opisali; naj povemo le, da sloni na MySQL in tehnologiji PHP), analitičnem strežniku, ki iz baze črpa podatke, jih primerja med seboj, gradi hierarhične modele skupin genov oziroma izračuna obogatitve genskih skupin v genski ontologiji, predlaga genske mreže ipd., ter aplikacije na odjemalcu, ki vse skupaj primerno pokaže uporabniku. V pričujočem prispevku smo se osredotočili na predstavitev funkcij grafičnega vmesnika in ga opisali s stališča uporabnika.

Aplikacija dictyEpress danes služi za prikaz genskih izraznih profilov že objavljenih podatkov in eksperimentov, a tudi – ob ustrezni avtentikaciji – za prikaz rezultatov še neobjavljenih eksperimentov. Ti so dostopni v spletni aplikaciji takoj, ko so vnešeni v bazo podatkov. DictyExpress je sicer samostojna aplikacija, ki pa je skupnosti raziskovalcev socialne amebe *Dictyostelium* na voljo tudi neposredno iz domače strani tega organizma v portalu dictyBase. Tako je na vsaki domači strani posameznega gena tega organizma povezava v dictyExpress, ki aplikacijo odpre tako, da se prikaže izrazni profil gena v divjem tipu. Seveda je dictyExpress na ta način popolnoma nadomestil prejšnje, statične izrise izraznih profilov na strani dictyBase.

## Zaključek

Spletna aplikacija dictyExpress služi sorazmerno majhni skupini nekaj več kot 1000 raziskovalcev organizmov *Dictyostelium*. Dnevno aplikacijo obiše okoli 5% te skupnosti, kar je pravzaprav kar izjemno. Na srečo za razvijalce to število omogoča izvajanje aplikacije na enem samem, sicer močnejšem strežniku, ki lahko podpre sočasno uporabo (v mejah predvidene odzivnosti) do deset uporabnikov. Od časa objave uradne aplikacije v pričetku jeseni 2009 do danes je bilo pozitivnih in pohvalnih odzivov uporabnikov veliko, na veselje razvijalcev pa precej manj zahtev po novih funkcijah. Kot vse kaže so najbolj zahtevni uporabniki sistema prav tisti, s katerimi smo aplikacijo razvili.

DictyExpress ni splošno uporabna, generična aplikacija s področja analitike izraznih profilov. Omejuje se na čisto določene funkcije in metode prikazov podatkov, na določene tipe uporabniških interakcij z uporabnikom, ter na določen tip (časovno odvisnih) podatkov. Te omejitve pa ne zmanjšujejo uporabnosti aplikacije. S stališča ciljnih uporabnikov prav nasprotno. Eden od glavnih adutov dictyExpressa je prav njegova enostavnost, ta pa ravno izhaja iz omejenega nabora funkcij, ki očitno je za ciljne uporabnike ravno pravšnja.

Aplikacija dictyExpress je razvita za okolje Flash, njegova interaktivnost pa temelji na uporabi odličnih grafičnih gradnikov, ki jih to okolje ponuja in ki posnemajo gradnike namiznih aplikacij. To je dobro seveda za vse uporabnike računalnikov, ki smo navajeni namiznih aplikacij. Alternative okolju Flash seveda obstajajo (JavaScript, HTML 5 idr.), so s stališča integracije z jezikom za opis spletnih strani HTML nekako "čistejše", a za razvijalce trenutno precej manj ugodne zaradi velikega pomanjkanja dobrih razvojnih orodij. Pričakujemo seveda, da se bo to v kratkem spremenilo, kar bomo seveda občutili razvijalci aplikacij. Uspešne bodo le, če uporabniki tehnoloških sprememb sploh ne bodo opazili.

## Zahvala

Projekt razvoja aplikacije dictyExpress je finančno podprla Javna agencija za raziskovalno dejavnost Republike Slovenije (P2-0209, J2-9699, L2-1112).

## Literatura

1. Rot G, Parikh A, Curk T, Kuspa A, Shaulsky G, and Zupan B. dictyExpress: a *Dictyostelium* discoideum gene expression database with an explorative data analysis web-based interface. *BMC Bioinformatics*. 2009; 10: 265.
2. Bonner JT. *The social amoebae: the biology of cellular slime molds*. Princeton University Press; 2009.
3. Van Driessche N, Shaw C, Katoh M, Morio T, Sugang R, Ibarra M, et al. A transcriptional profile of multicellular development in *Dictyostelium discoideum*. *Development*. 2002, 129(7):1543-52.
4. Fey P, Gaudet P, Curk T, Zupan B, Just EM, Basu S, et al. dictyBase – a *Dictyostelium* bioinformatics resource update. *Nucleic Acids Res*. 2009, 37(Database issue):D515-9.
5. Van Driessche N, Demsar J, Booth EO, Hill P, Juvan P, Zupan B, Kuspa A, Shaulsky G. Epistasis analysis with global transcriptional phenotypes. *Nat Genet*. 2005, 37(5):471-7.
6. Gene Ontology Consortium. The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res*. 2010, 38(Database issue):D331-5.