

آنالیز کلاستر الگویی بیان ژنهای پاسخگو به شوری

حسین عسکری^۱، محمد کافی^۲، قاسم حسینی سالکده^۳

۱. پژوهشکده بیوتکنولوژی کشاورزی، h_askari@abrii.ac.ir

۲. دانشگاه فردوسی مشهد، kafi-m@ferdowsi.um.ac.ir

۳. پژوهشکده بیوتکنولوژی کشاورزی، h_salekdeh@abrii.ac.ir

چکیده:

هم تکاملی فرایندی تکاملی است در طی آن تغییراتی قابل توارث در یک نهاد منجر به ایجاد یک فشار انتخابگر برای تغییر در نهادها می شود. هم تکاملی بیان در مقایسه با هم تکاملی توالی اسیدهای آمینه می تواند به عنوان یک پیشگویی قوی از برهمکنش های فیزیکی باشد (Fraser et al., 2004). به منظور بدست آوردن سربهایی از ژنهای هم تکامل روشهای متعدد آنالیز کلاستر ایجاد شده است که بر پایه آن می توان ژنهای هم بیان را از داده های ریزآرایه جدا کرد اما تاکنون مطالعات اندکی بر روی داده های پروتئوم اجرا شده است. در مطالعه حاضر، روشهای آنالیز کلاستر شامل Adaptive quality-based clustering، Fuzzy Adaptive Resonance Theory (FuzzyART)، Self Organizing Map (SOM)، Hierarchical clustering و Adaptive quality-based clustering مورد استفاده قرار گرفت. نتایج این پژوهش نشان داد که FuzzyART بالاترین توانایی را در مقایسه با Hierarchical و SOM برای کلاستر ۱۰۲ پروتئین پاسخگو به شوری در ۱۶ گروه داشت. روش Adaptive quality-based clustering دو گروه مشخص با ۴۱ و ۱۱ عضو را ایجاد کرد. بر پایه این روش، اعضا هر کلاستر با سطح اطمینان ۹۵٪ تحت شرایط این آزمایش هم بیان هستند. نتایج ما نشان داد که کاربرد دو روش آنالیز کلاستر FuzzyART و Adaptive quality-based clustering که دارای کارکردی متفاوت و مکمل هستند می تواند منجر به کلاستر شدن پروتئین های هم بیان در گروه های مشابه گردد.

Abstract:

Coevolution is an evolutionary process in which a heritable change in one entity establishes selective pressure for a change in another entity. Coevolution of expression can be more powerful predictor of physical interaction than is coevolution of amino acid sequence (Fraser et al. 2004). In order to gain series of coevolved genes in a given gene population, several clustering techniques have been developed to discover coexpressed genes from microarray data. However, clustering of proteome data has not been well studied yet. In this study, we applied some clustering techniques (Hierarchical clustering, SOM- Self Organizing Map, FuzzyART- Fuzzy Adaptive Resonance Theory, and Adaptive quality-based clustering) to cluster coexpressed salt response-related proteins from *suaeda* plants. The results revealed that FuzzyART had the highest ability to cluster 102 salt responsive proteins in 12 clusters compared to Hierarchical and SOM methods. Adaptive quality-based clustering algorithm made 2 remarkable clusters with 41 and 11 members. With confidence level of 95%. Our results showed that application of FuzzyART and Adaptive quality-based clustering, two clustering methods with different and complementary functions, can successfully cluster coexpressed proteins.

مقدمه:

پیشرفت سریع در توالی یابی ژنوم، موجب گسترش روشهای کاوش این اطلاعات با تکیه بر تشریح نوین فرایندهای بیولوژیک گردیده است (Eisen et al., 1998). تکنیکهای ریز آرایه و آنالیز پروتئوم که امکان بررسی همزمان بیان چند صد تا چند هزار ژن را فراهم می نمایند حجم عظیمی از اطلاعات بیولوژیک را ایجاد کرده اند که جدای از شناخت بیان تمایزی ژنها ضرورت تفسیر کارکردی (Functional annotation) ژنهای هم بیان (Co-expressed genes) را بیش از پیش مورد نظر قرار می دهد. در این رابطه قدم نخست به طبقه بندی ژنهایی با الگوی بیان مشابه باز می گردد (Sturn et al., 2002). استفاده از روشهای کمی کلاستر در سریهای متوالی تیمارهای آزمایشی در رابطه با بیان پروتئین ها کمتر مورد استفاده قرار گرفته است (Lonosky et al., 2004). این پژوهش به منظور ارزیابی روشهای کمی کلاستر جهت تعیین پروتئینهای هم بیان پاسخگو به تیمارهای شوری در گیاه *Suaeda* اجرا شده است.

مواد و روشها:

گیاه *Suaeda aegyptiaca* به مدت ۳۰ روز در معرض ۵ سطح از تیمار شوری (۰، ۱۵۰، ۳۰۰، ۴۵۰ و ۶۰۰ میلی مولار سدیم کلراید) قرار گرفت. کل پروتئین های محلول در برگ گیاه با استفاده از روش الکتروفورز دو بعدی تجزیه شدند. استخراج پروتئین، اجرای الکتروفورز دو بعدی و آنالیز ژل مطابق Hajheidari و همکاران (۲۰۰۵) انجام شد. از میان ۷۰۰ نقطه پروتئینی آنالیز شده، ۱۰۲ پروتئین پاسخ معنی داری به شوری نشان دادند. معیار معنی دار بودن تأثیر شوری بر بیان پروتئین بر پایه دو شاخص (تغییر بیان به میزان حداقل ۲ برابر نسبت به شاهد و یا معنی دار شدن تغییر بیان در ۳ سطح آماری با شاخص $LSD, p \leq 0.05$) تعیین شد. روشهای آنالیز کلاستر مورد بررسی عبارت بودند از Hierarchical clustering (Eisen et al., 1998)، Self Organizing Map (SOM- Tamayo et al., 1999)، Fuzzy Adaptive Resonance Theory (FuzzyART-Tomida et al., 2002) و Adaptive quality-based clustering (De Smet et al., 2002). قواعدی که در این پژوهش برای ارزیابی روش های آنالیز کلاستر استفاده گردیده است بر اساس دو روش محاسباتی میانگین انحراف استاندارد (ASD- Averag Standard Deviation) و شاخص اختلاف سطح بود (Gap area index) (Tomaid et al., 2002). که اولی برای محاسبه بهترین تعداد کلاستر و دومی برای مقایسه دقت روشهای آنالیز استفاده شد.

نتایج و بحث:

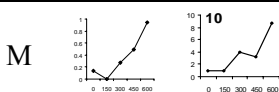
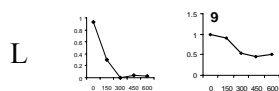
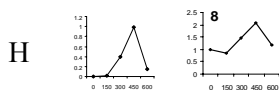
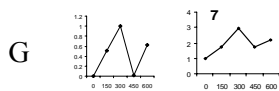
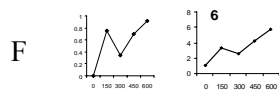
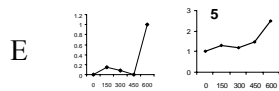
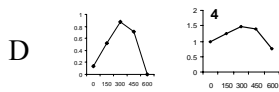
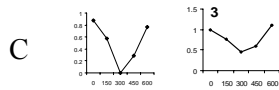
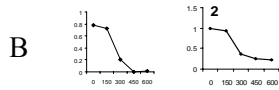
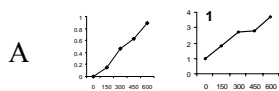
آنالیز کلاستر الگویی بیان عمدتاً با موانعی از جمله خطای آنالیز کلاستر (Clustering error)، افزایش بی مورد تعداد گروه ها (Over clustering) و عدم تطابق گروه ها با مفاهیم بیولوژیک (Biological invalidation of clustering) روبرو هستند. توسعه و انتخاب بهترین روشهای آنالیز کلاستر بر پایه طبیعت داده ها یکی از بخشهای مهم ارزیابی حجم عظیم داده های حاصل از روشهای پر بازده (high throughput) مطالعه بیان ژن است. تعیین بهترین تعداد کلاستر قدم نخست در آنالیز کلاستر خواهد بود. Hierarchical clustering یکی از مرسوم ترین روشهای آنالیز کلاستر است که دارای دو ویژگی خاص یعنی تعیین روابط خویشاوندی بین الگوهای بیان ژن و نمایش رنگی از روند بیان ژنها می باشد. دو روش SOM و FuzzyART از جمله روشهای مبتنی بر اصول شبکه های عصبی است. در روش SOM تعیین تعداد

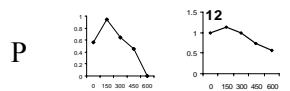
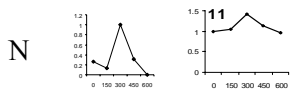
گروه به عنوان پیش فرض در نظر گرفته شده است در حالی که در روش FuzzyART تعداد گروه به عنوان پیش فرض مطرح نیست چگونه ای که در این الگوریتم از پارامتری با عنوان Vigilance استفاده می شود که معیاری ($0 \leq p \leq 1$) از شباهت بین الگوی بیان یک ژن و پروفیل هر کلاستر است. برای تعیین بهترین تعداد کلاستر، ۱۰۲ پروتئین انتخاب شده با روش برای بدست آوردن تعداد ۹، ۱۰، ۱۱، ۱۲، ۱۳ و ۱۴ کلاستر مورد آزمون قرار گرفتند (شکل ۱). این نتایج نشان داد کمترین ASD و در نتیجه بهترین تعداد کلاستر زمانی بدست آمد که ۱۰۲ پروتئین در ۱۲ کلاستر طبقه بندی شوند. این در حالیست که تغییر تعداد کلاستر در روش SOM تاثیر قابل ملاحظه ای بر روند کاهش ASD نداشت. در مرحله بعد، روشهای آنالیز کلاستر با استفاده از شاخص اختلاف سطح و بر مبنای تعداد کلاستر مساوی (۱۲ عدد) برای هر ۳ روش مورد ارزیابی قرار گرفت. نتایج این ارزیابی که با تکیه بر محاسبه اختلاف سطح زیر منحنی ایجاد شده بین پروفیل هر عضو یک کلاستر با پروفیل میانگین همان کلاستر بدست آمده است نشان داد روش FuzzyART قادر به ایجاد حداقل شاخص اختلاف سطح می باشد (شکل ۲). ارزیابی پروفیل کلاسترهای ایجاد شده توسط ۳ روش مذکور نشان داد که پروفیل‌های حاصل از روشهای Hierarchical و FuzzyART بیشترین شباهت را به یکدیگر دارند (شکل ۳). یکی دیگر از روشهای مورد استفاده در آنالیز داده های حاصل از ریزآرایه (Microarray) روش Adaptive quality-based clustering است که بر اساس تعیین احتمال هم بیانی ژنهای مورد آنالیز عمل می نماید. در این روش بخشی از ژنها که در سطح احتمال مورد بررسی در کلاستر قرار نمی گیرند از مراحل بعدی آنالیز حذف می شوند. نتایج این بررسی نشان داد که از ۱۰۲ پروتئین ورودی، ۵۲ پروتئین در دو کلاستر شامل ۴۱ و ۱۱ عضو قرار گرفتند. این بررسی بیان می نماید که اعضا قرار گرفته در هر کلاستر با احتمال ۹۵٪ با یکدیگر تنظیم بیان می شوند و احتمالاً بر این عملکرد آنها در تعیین سطح تحمل گیاه نقش کلیدی دارد (شکل ۴). این نتایج همچنین نشان داد برای مقایسه توانایی روشهای آنالیز کلاستر، استفاده از روشهای کمی بخوبی کاربرد دارد این در حالیست که Lonosky و همکاران (۲۰۰۴) بر پایه یک مقایسه کیفی روشهای آنالیز کلاستر را با یکدیگر مقایسه کرده اند. نتایج بدست آمده از مطالعات Tomida و همکاران (۲۰۰۲) بر روی داده های ریزآرایه نشان داد که روش FuzzyART قدرت بالاتری در شناسایی الگوی بیان مشابه دارد و این بخوبی برای آنالیز کلاستر پروتئین های پاسخگو به شوری قابل استفاده است. استفاده از روشهایی همچون Adaptive quality-based clustering قادر است به عنوان یک روش مکمل در کنار روش هایی همچون FuzzyART نتایج مناسبی در خصوص احتمال هم بیانی پروتئین ها ارائه دهد.

منابع

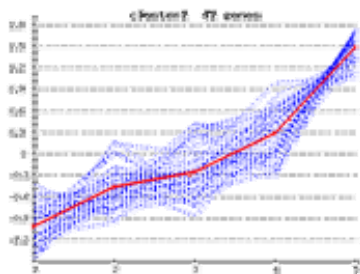
1. Sturn, A., J. Quackenbush, and Z. Trajanoski. 2002. *Bioinformatics* 18: 207-208
2. Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. *PNAS* 95: 14863-14868
3. Tomida, S., T. Hanai, H. Honda, and T. Kobayashi. 2002. *Bioinformatics* 18: 1073-1083
4. Lonosky, P.M., X. Zhang, V.G. Honavar, D.L. Dobbs, A. Fu, and S.R. Rodermel. 2004. *Plant Physiol.* 134: 560-574
5. Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub. 1999. *PNAS* 2907-2912
6. De Smet, F., J. Mathys, K. Marchal, G. Thijs, B. De Moor, and Y. Moreau. 2002. *Bioinformatics* 18: 735-746
7. Hajheidari, M., M.A. Noghabi, H. Askari, M. Heidari, S.Y. Sadeghian, E.S. Ober, and G.H. Salekdeh. 2005. *Proteome analysis of sugar beet leaves under drought stress*. 5: 950-960

FuzzyART based
 on
 Name _____
 weight average
 verctor profile

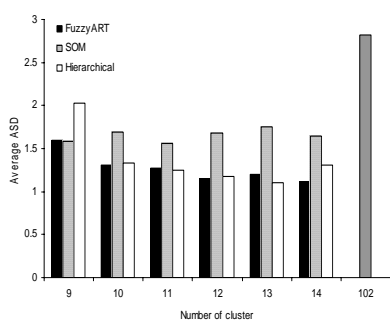




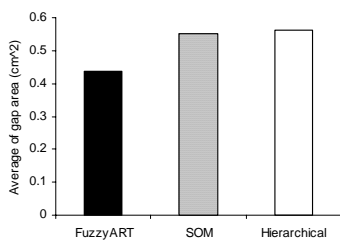
جدول ۱: الگوهای حاصل از آنالیز کلاستر بیان ژن با استفاده از روش FuzzyART



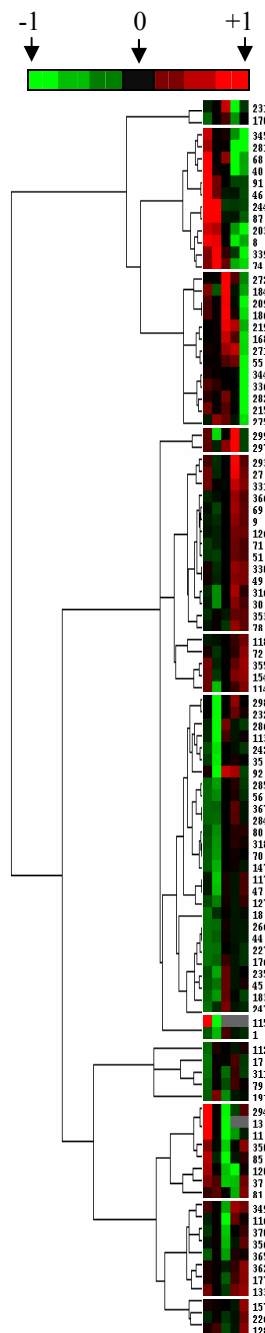
شکل ۱: الگوی شجره ای حاصل از آنالیز کلاستر با روش Hierarchical clustering



شکل ۲: تعیین تعداد بهینه کلاستر با استفاده از ASD



شکل ۳: مقایسه توانایی روشهای مختلف آنالیز کلاستر در ایجاد ۱۲ کلاستر



شکل ۴: کلاسترهای ایجاد شده در روش Adaptive quality-based clustering. محور افقی، سطوح مختلف شوری و محور عمودی داده های نرمال شده از الگوی بیان ژن.