Ferdowsi University of Mashhad

ISSN 2008-9147
Numbers: 18

# JCMR

## Journal of Cell and Molecular Research

Volume 9 , Number 2, Winter 2017

بسم الله الرحمن الرحيم

Issuance License No. 124/902-27.05.2008 from Ministry of Culture and Islamic Guidance
Scientific Research Issuance License No. 161675 from the Ministry of Science, Research and Technology, Iran

# Journal of Cell and Molecular Research (JCMR)

**Volume 9, Number 2, Winter 2017**

**Copyright and Publisher**
*Ferdowsi University of Mashhad*



---

**JCMR Office:** Department of Biology, Faculty of Sciences, Ferdowsi University of Mashhad, Mashhad, Iran.
**Postal Code:** 9177948953
**P.O. Box:** 917751436
**Tel:** +98-513-8804063
**Fax:** +98-513-8795162
**E-mail:** jcmr@um.ac.ir
**Online Submission:** http://jcmr.fum.ac

# Table of Contents

*Journal of Cell and Molecular Research (2017) 9 (2), 54-58* DOI: 10.22067/jcmr.v9i2.67315

Research Article

# Karyosystematics of an Endemic tooth-carp, *Aphanius shirini* (Teleostei: cyprinodontidae) from Iran

Azam Mansoori[1], Mehregan Ebrahimi[1,2], Ali Gholamhosseini[1], Hamid Reza Esmaeili[1,*]

[1] *Department of Biology, College of Sciences, Shiraz University, Shiraz, Iran*
[2]*School of Biological Sciences, Flinders University, Adelaide, SA, Australia*

Received 8 Septamber 2017 Accepted 15 October 2017

## Abstract

The karyological and cytological characteristics of an endemic cyprinodont fish of Iran, *Aphanius shirini* have been investigated for the first time by examining metaphase chromosomes spreads obtained from gill epithelial and kidney cells. The diploid chromosome number of this species is 48. The karyotype consisted of one submetacentric and 23 subtelocentric pairs of chromosomes (2Sm + 46St). The chromosome arm number (NF) is 50. Sex chromosomes were cytologically indistinguishable in this tooth-carp. Based on the present and previous reported diploid chromosome number for other cyprinodont species, it can be suggested that the diploid chromosome number of 2n = 48 is the modal number of the cyprinodont fish.

***Keywords*:** Cyprinodontiformes, Chromosome, Cytogenetical analysis, Idiogram

## Introduction

Cyprinodontiformes order comprises 10 families and about 1326 species (Eschmeyer and Fong, 1998) of mostly small, fresh and brackish water fish inhabiting harsh environments, such as saline or very warm waters, water of poor quality, or isolated situations where no other types of fish live (Esmaeili et al., 2016; Gholami et al., 2014). The Cyprinodontidae are represented in Iran by only one genus *Aphanius* Nardo, 1827. From a total of 32 *Aphanius* species which have been described around the world, one fossil record, *Aphanius persicus* and 14 species have been reported from Iranian drainages: *A. arakensis* Teimori, Esmaeili, Gholami, Zarei and Reichenbacher, 2012 ; *A. darabensis* Esmaeili, Teimori, Gholami and Reichenbacher, 2014; *A. dispar* Rüppell, 1829; *A. farsicus* Teimori, Esmaeili and Reichenbacher, 2011; *A. furcatus* Teimori, Esmaeili, Erpenbeck and Reichenbacher, 2014; *A. ginaonis* Holly, 1929; *A. isfahanensis* Hrbek, Keivany and Coad, 2006; *A. kavirensis* Esmaeili, Teimori, Gholami and Reichenbacher, 2014; *A. mento* Heckel, 1843; *A. mesopotamicus* Coad, 2009; *A. pluristriatus* Jenkins, 1910; *A. shirini* Gholami, Esmaeili, Erpenbeck and Reichenbacher, 2013; *A. sophiae* Heckel, 1847 and *Aphanius vladykovi* Coad, 1988, (Jouladeh-Roudbar et al., 2015). Till now, the karyological studies of seven *Aphanius* species (out of 14 described species) have been reported from Iran consisted of *A. dispar* and *A. ginaonis* (Esmaeili et al., 2008a), *A. farsicus* and *A. sophiae* (Esmaeili et al., 2007), *A. isfahanensis* (Esmaeili et al., 2008b), *A. mento* (Arai, 2011) and *Aphanius vladykovi* (Esmaeili et al., 2009). *Aphanius shirini*, *Gholami*, Esmaeili, Erpenbeck and Reichenbacher, 2014 or Kapour-e-dandandar-e-Khosroshirin (Farsi); Shirin or Khosroshirin tooth-carp (English) and Khosroshirin Zahnkärpflinge (German) is an endemic species found in the uppermost reaches of the Kor River Basin. Khosroshirin tooth-carp is distinguished from other Iranian species of *Aphanius* by having the lowest number of flank bars among the Iranian inland *Aphanius* species, molecular characters of mitochondrial cytochrome b, DNA sequence data and multivariable morphometric and meristic traits. Thus, the Khosroshirin population clearly represents a new species based on both molecular and morphological evidence (Gholami et al., 2014).

Tooth-carps of Iran have been studied mainly based on their morphology but species identification on this basis is not always possible. The application of non-morphological methods such as cytogenetic studies may provide a complementary data source for more accurate and precise identification of these fishes. Fish karyosystematics is a branch of systematics that links systematics, cytology, and

Corresponding authors E-mail:
*hresmaeili@shirazu.ac.ir

genetics to find out structure and evolution of karyotypes and to reconstruct phylogenetic relationship of fish taxa (Yu et al., 1987). Application of this type of studies has received considerable attention in recent years (Esmaeili and Shiva, 2006; Galetti Jr et al., 2000; Harrison et al., 2007). Fish chromosome data have great importance in studies concerning evolutionary systematics, aquaculture, mutagenesis, genetic control and the rapid production of inbred lines (Al-Sabti, 1991). The increasing importance of chromosomal studies on fish and lack of data on karyotyping of Khosroshirin tooth-carp encouraged us to do first cytogenetical analysis (i.e., diploid chromosome number, description of karyotype, idiogram) of this endemic tooth-carp of Iran.

## Materials and Methods

*Aphanius shirini* specimens were collected from the Paselari spring of the Khosroshirin spring-stream system, uppermost reaches of Kor River basin, Khosroshirin Village, Abadeh City, Fars, Iran, 30° 53′29.5″ N 52° 00′36.8″ E, Alt. 2327 m (Figure 1) using a dip net. The fishes were transported live to the laboratory, and kept in a well-aerated aquarium at 20 – 25°C before analysis. For karyological studies, the modified method of Uwa (1986) was used. Vinblastine solution was prepared with 0.005 g in 20 ml of physiological serum. The fish were injected intraperitoneally with 0.02 ml of vinblastine per gram of body weight using an insulin syringe and then were put back in the aquarium for 3 - 4 h.

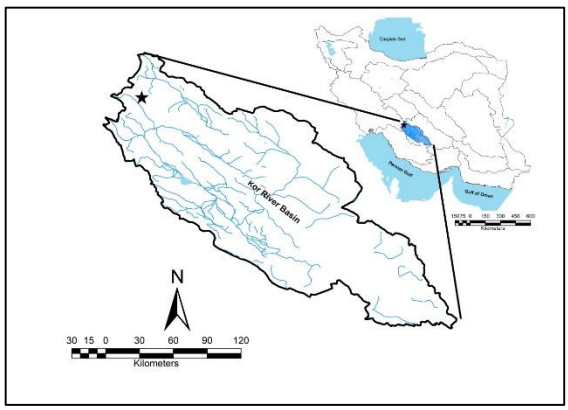



**Figure 1.** Location of *Aphanius shirini* population analyzed in this study.

The gill filaments and kidneys of those specimens were then removed and placed in hypotonic 0.36% KCI solution for 45 min at room temperature (25°C). Thereafter, the solutions were centrifuged for 10 min at 1000 rpm, adding 2 - 3 drops of fresh and cold Carnoy's fixative (1 : 3, acetic acid: methanol) before centrifugation. The supernatants were then discarded and 5 ml of fresh and cold fixative was added to the sediments, which were mixed thoroughly and then left for 1 h. The fixation and centrifugation stages were repeated twice. The suspensions were then trickled onto cold slides. These slides were stained with 20% Giemsa for 20 min. Chromosomes were observed, selected and photographed by Nikon light microscope with a camera mounted on it. Karyotypes were prepared by arranging chromosomes in pairs by size. For each chromosome, the average lengths of the short and long arms and arm ratio (the ratio of the long arm length to the short arm length of chromosomes) were calculated and then the chromosomes were classified according to the criteria given by Levan et al. (1964). Fundamental number (NF) was expressed as twice the number of atelocentric chromosomes plus the number of telocentric chromosomes. The idiogram was prepared in Harvard Graphics 2.0 software.

## Results

Metaphase spread of this species is given in Figure 2. The diploid chromosome number was 2n = 48 (Figure 3).

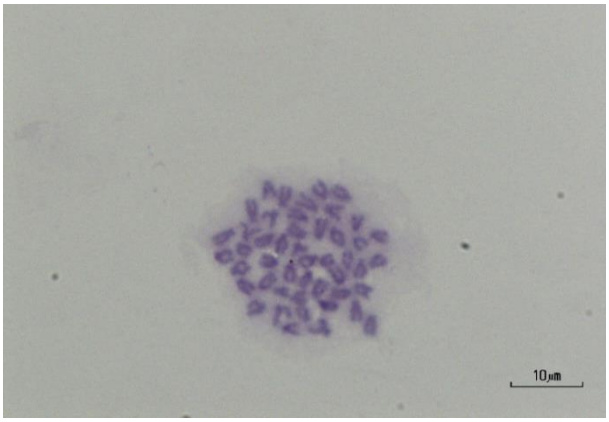



**Figure 2.** Giemsa stained metaphase chromosome spread of Aphanius shirini from Iran.

**Figure 3**. Giemsa stained karyogram of *Aphanius shirini* from Iran.

The quantitative data of the different measurements used to classify chromosomes and the idiogram are given in Table 1 and Figure 4 respectively. The karyotype consisted of one pair of submetacentric and 23 pairs of subtelocentric chromosomes (2Sm + 46St), and the arm number (NF) was 50. Sex chromosomes were cytologically indistinguishable in this endemic tooth-carp.

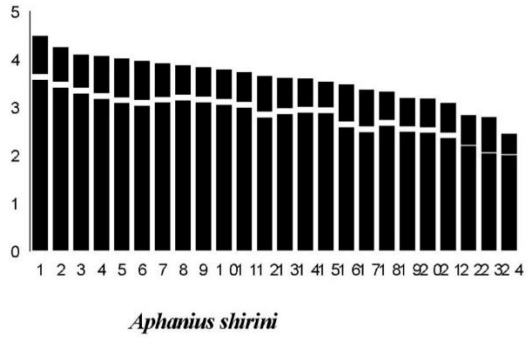



**Figure 4.** Haploid idiogram of *A. shirini* from Iran.

**Table 1**. Chromosome measurements (in µm) and classification of *Aphanius shirini* chromosomes (Ch. No.: Chromosome number; LA: Long arm; SA: Short arm; TL: Total length; AR: Arm ratio; CT: Chromosome type; Sm: Submetacentric; St: Subtelocentric).

| Ch. No. | LA | SA | TL | AR | CT |
|---|---|---|---|---|---|
| **1** | 3.58 | 0.81 | 4.39 | 4.37 | St |
| **2** | 3.42 | 0.74 | 4.17 | 4.58 | St |
| **3** | 3.30 | 0.71 | 4.01 | 4.63 | St |
| **4** | 3.18 | 0.79 | 3.98 | 4.00 | St |
| **5** | 3.10 | 0.83 | 3.93 | 3.70 | St |
| **6** | 3.04 | 0.83 | 3.88 | 3.63 | St |
| **7** | 3.11 | 0.72 | 3.83 | 4.29 | St |
| **8** | 3.15 | 0.63 | 3.79 | 4.97 | St |
| **9** | 3.11 | 0.63 | 3.75 | 4.89 | St |
| **10** | 3.06 | 0.63 | 3.69 | 4.83 | St |
| **11** | 2.99 | 0.65 | 3.64 | 4.61 | St |
| **12** | 2.79 | 0.77 | 3.56 | 3.63 | St |
| **13** | 2.87 | 0.65 | 3.52 | 4.39 | St |
| **14** | 2.90 | 0.61 | 3.51 | 4.74 | St |
| **15** | 2.89 | 0.54 | 3.43 | 5.30 | St |
| **16** | 2.59 | 0.79 | 3.39 | 3.26 | St |
| **17** | 2.49 | 0.78 | 3.27 | 3.17 | St |
| **18** | 2.62 | 0.61 | 3.24 | 4.28 | St |
| **19** | 2.49 | 0.60 | 3.10 | 4.14 | St |
| **20** | 2.47 | 0.61 | 3.09 | 4.00 | St |
| **21** | 2.36 | 0.63 | 3.00 | 3.71 | St |
| **22** | 2.20 | 0.63 | 2.84 | 3.45 | St |
| **23** | 2.05 | 0.75 | 2.80 | 2.70 | Sm |
| **24** | 2.01 | 0.44 | 2.46 | 4.48 | St |

## Discussion and Conclusion

According to our observations, the diploid chromosome number of *Aphanius shirini* species was 2n = 48 and is in conformation with the chromosome number of other species of this genus. The chromosome numbers of studied *Aphanius* species including *A. sophiae*, *A. farsicus*, *A. asquamatus*, *A. dispar*, *A. fasciatus*, *A. iberus* and *A. mento* have been reported to be 2n = 48. Hence, it can be concluded that the chromosome number in this genus is conserved. The number of chromosomes in this tooth-carp is also similar to that of other species of Cyprinodontidae such as *Cyprinodon alvarezi*, *Cyprinodon atrorus* and *Cyprinodon beltrani*. In the order Cyrinodontiformes, the most common fish species which have so far been cytologically investigated, such as *Gambusia affinis*, *G. holbrooki*, *G. gaigei*, *G. nobilis*, *Girardinus metallicus*; *Poecilia vivipara* (Poecillidae); *Fundulus diaphanus* (Fundulidae); *Allotoca maculata*, *Goodea luitpoldi*, *G. atripinnis*, *G. gracilis*, *Hubbsina turneri*, *Ilyodon furcidens*, *Ilyodon lennoni*, *Skiffia francesae*, *Skiffia bilineata*, *Xenoophorus captivus*, *Xenotaenia resolanae*, *Xenotoca eiseni*, *X. melanosoma*, *X. variata* (Goodeidae), have the diploid chromosome number of 2n = 48 (Arai, 2011). Yet in a few species of Cyprinodontiformes such as *Aphyosemion bivittatum*, *A. bualanum*, *A. calliurum*, *Fundulopanchax sjostedti*, *Fundulopanchax mirabilis* (Aplocheilidae); *Allotoca dugesi* ,*Allodontichthys hubbsi* and *Ameca splendens* (Goodeidae) the diploid chromosome number is reported to vary from 2n = 26 to 2n = 42 (Arai, 2011). It could be suggested that the diploid chromosome number of 2n = 48 is the modal number of cyprinodont fish. In the interpretation of karyotypic evolution, it is often assumed that the primitive fish karyotype consists of 48 rods from which the karyotypes of all existing fish forms have been derived (Khuda-Bukhsh et al., 1986) but the issue seems yet to be resolved. The discovery of 48 rather large acrocentric chromosomes in the Pacific hagfish, *Eptatretus stoutii*, belonging to the order Myxiniformes (Taylor, 1967; Vasil'yev, 1980) and the occurrence of 48 rods in the majority of fishes studied prior to 1967 led to the idea that the primitive karyotype of ancestral vertebrate freshly evolved from chordate might consist of 48 rods (Khuda-Bukhsh et al., 1986). Therefore, most of the subsequent workers assumed the karyotypic evolution in different groups of fishes based on this basic assumption of 48 rods as the primitive number (Khuda-Bukhsh et al., 1986). But the discovery of 2n = 24 rods in two species of freshwater eels (Kitada and Tagawa, 1973; Rishi and Haobam, 1984), 2n = 36 rods in two species of *Myxine*, low diploid numbers ranging between 14 and 42 in a large

number of fish families showing NF less than 36 in some cases (Khuda-Bukhsh et al., 1986) would possibly call for a more cautious prediction on the primitive karyotype of fish.

In the present study, no cytological evidence was found for sex chromosome dimorphism which agrees with reports on many fish species such as Serranidae and Mugilidae (Aguilar 1997, Rossi *et al*. 1997).

The karyotype formula of this tooth-carp was 2Sm + 46St and the chromosome arm number was 50. Chromosome formula of 16Sm + 32St was reported for *A. dispar* and *A. farsicus*; 14Sm + 34St for *Aphanius ginaonis*; 12Sm + 34St in *A. isfahanensis* and 8Sm + 40St for *A. sophiae* and *Aphanius vladykovi*. The arm number of NF = 32 was reported for *A. dispar* and *A. farsicus* and NF = 28 for *Aphanius sophiae* and *A. vladykovi*. The arm number in *Aphanius ginaonis* and *A. isfahanensis* were reported to be 31 and 30 respectively (Esmaeili et al., 2008a; Esmaeili et al., 2008b, 2009; Esmaeili et al., 2007). Though chromosome numbers of *Aphanius* species are conserved despite different geographical locations, the fundamental arm numbers are different. These differences within *Aphanius* species of different geographical locations suggest that structural rearrangement in chromosome complements, as a consequence change in chromosome morphology without any change in chromosome number. This divergence may be attributed to differences in the karyotype macrostructure, reflecting a real geographical variation common to widespread species or may be the result of differences in the scoring of submetacentric or metacentric chromosomes as different degrees of chromosome condensation, leading to differences in chromosome classification.

**Acknowledgements**

The authors give special thanks to Amin Marashi for his help during field and labratory works. Financial support was provided by Shiraz University which is gratefully acknowledged.

**Conflicts of Interest**

The authors declare that there is no conflict of interest regarding the publication of this article.

**References**

1. Al-Sabti K. (1991) Handbook of genotoxic Effects and fish chromosomes. Joseph Stefan Institute, Ljubljana, Slovenia.
2. Arai R. (2011) Fish Karyotypes: A Check list. Springer science and business media, Japan.
3. Eschmeyer W. and Fong J. (1998) Catalog of fishes. *In* Species by Family/Subfamily. Vol. 2016, San Francisco, USA.
4. Esmaeili H. R., Ebrahimi M. and Saifali M. (2008a) Karyological analysis of five tooth-carps (Actinopterygii: Cyprinodontidae) from Iran. Micron 39:95-100.
5. Esmaeili H. R., Ebrahimi M., Teimory A. and Ansary T. H. (2008b) First karyological analysis of an endemic fish, Isfahan tooth-carp, *Aphanius isfahanensis* (Actinopterygii: Cyprinodontidae) from Iran. Journal of Applied Animal Research 33:73-76.
6. Esmaeili H. R., Ebrahimi M., Teimory A. and Ansary T. H. (2009) First karyological analysis of an endemic fish, Zagros tooth-carp, *Aphanius vladykovi* Coad, 1988 (Actinopterygii: Cyprinodontidae) from Iran. Iranian Journal of Science and Technology (Sciences) 33:349-354.
7. Esmaeili H. R., Masoudi M., Ebrahimi M. and Elmi A. (2016) Review of *Aphanius farsicus*: a critically endangered species (Teleostei: Cyprinodontidae) in Iran. Iranian Journal of Ichthyology 3:1-18.
8. Esmaeili H. R., Piravar Z. and Shiva A. (2007) Karyological analysis of two endemic tooth-carps, *Aphanius persicus* and *Aphanius sophiae* (Pisces: Cyprinodontidae), from Southwest Iran. Turkish Journal of Zoology 31:69-74.
9. Esmaeili H. R. and Shiva A. H. (2006) Reproductive biology of the Persian tooth-carp, *Aphanius persicus* (Jenkins, 1910)(Cyprinodontidae), in southern Iran. Zoology in the Middle East 37:39-46.
10. Galetti Jr P., Aguilar C. and Molina W. (2000) An overview of marine fish cytogenetics. *In* Marine Genetics. A.M. Sole-Cava, C.A.M. Russo, and J.P. Thorpe, editors. Springer, Dordrecht, Netherlands 55-62.
11. Gholami Z., Esmaeili H. R., Erpenbeck D. and Reichenbacher B. (2014) Phylogenetic analysis of *Aphanius* from the endorheic Kor River Basin in the Zagros Mountains, South-western Iran (Teleostei: Cyprinodontiformes: Cyprinodontidae). Journal of Zoological Systematics and Evolutionary Research 52:130-141.
12. Harrison I., Nirchio M., Oliveira C., Ron E. and Gaviria J. (2007) A new species of mullet (Teleostei: Mugilidae) from Venezuela, with a discussion on the taxonomy of *Mugil gaimardianus*. Journal of Fish Biology 71:76-97.
13. Jouladeh-Roudbar A., Vatandoust S., Eagderi S., Jafari-Kenari S. and Mousavi-Sabet H. (2015) Freshwater fishes of Iran; an updated checklist. Aquaculture, Aquarium, Conservation and Legislation-International Journal of the Bioflux Society (AACL Bioflux) 8:892-894.
14. Khuda-Bukhsh A., Chanda T. and Barat A. (1986) Karyomorphology and evolution in some Indian hillstream fishes with particular reference

to polyploidy in some species. *In* Indo-Pacific Fish Biology. T. Uyeno, R. Arai, T. Taniuchi, and K. Matsuura, editors. Ichthyological Society of Japan, Tokyo, Japan 886-898.

15. Kitada J. and Tagawa M. (1973) On the chromosomes of two species of Cyclostomata. La Kromosomo 91:2913-2916.
16. Levan A., Fregga K. and Sandbeg A. A. (1964) Nomenclature for centromeric position on chromosomes. Hereditas 52:201-220.
17. Rishi K. and Haobam M. (1984) Karyotypic study on two freshwater mud-eels. *In* 4th All Indian Congress of Genetics. G.K. Manna and U. Sinha, editors. Hindasia, New Delhi, India 429-432.
18. Taylor K. M. (1967) The chromosomes of some lower chordates. Chromosoma 21:181-188.
19. Vasil'yev V. (1980) Chromosome numbers in fish-like vertebrates and fish. Journal of Ichthyology 20:1-38.
20. Yu X., Zhou T., Li K., Li Y. and Zhou M. (1987) On the karyosystematics of cyprinid fishes and a summary of fish chromosome studies in China. Genetica 72:225-235.

*Journal of Cell and Molecular Research (2017) 9 (2), 59-66* DOI: 10.22067/jcmr.v9i2.59568

Research Article

# Karyological Data of *Tanacetum polycephalum* Schultz-bip. and *T. parthenium* Schultz-bip. (Asteraceae) Populations

Hamideh Javadi *

*Research Institute of Forests and Rangelands, Tehran, Iran*

Received 16 October 2016 Accepted 15 November 2016

## Abstract

Chromosome numbers in 19 populations of *Tanacetum polycephalum* and *Tanacetum parthenium* from natural resources gene bank, that collected from different regions of Iran, were determined for the first time. The samples prepared by using root tips. After pretreatment, fixation, hydrolysis and staining, the microscopic samples prepared by squash method, metaphases were captured using an optical microscope. The best metaphases plates were selected and used for karyotype analyses. In all of populations the basic chromosome number was x=9 and the populations showed two ploidy levels (diploid & tetraploid). The type of the most chromosomes in all of the populations was metacentric (m) and sub-metacentric (sm) and located in 2A and 2B except for *T. parthenium* (Yazd, Taft) with 5m+3sm+1st karyotype formula and 2C Stebbins classes. In addition, *T. polycephalum* (Esfahan, Golpayegan) with the highest value of AR and $A_1$ had karyotype heterogeneity, also *T. polycephalum* (Esfahan, Golpayegan) and *T. polycephalum* (West-Azerbaijan, Uromeyeh) had the highest value of chromosome length (TL). Detailed karyotype allows us to group the different populations based on Stebbins classes and asymmetry indices.

***Keywords*:** Chromosome numbers, Ploidy levels, Karyotype, *Tanacetum polycephalum*, *Tanacetum parthenium*, Microscopic samples, Squash method

## Introduction

*Tanacetum* L. is a genus of about 160 species of flowering plants in the Aster family, Asteraceae native to many regions of the Northern Hemisphere (Watson, 1754), Northern Europe, Canada, Alaska, and Northern Russia (Heywood and Humphries, 1977; Tutin et al., 1976; Hulten, 1968), though the center of diversity for *Tanacetum* is South-West Asia and the Caucasus in the Old World (Soreng and Cope, 1991; Heywood and Humphries, 1977).

*Tanacetum* L. is a medicinal herb which is found in many old Gardens (Mitich, 1992). The genus *Tanacetum* (Asteraceae) is represented by 26 species in the flora of Iran, as herbal, perennial and sometimes shrub plants that dispersed in many regions of Iran, 12 of them are endemic (Mozaffarian, 2006). These species have traditionally been used as a spicy additive for food. It has been used in folk medicine for reducing fever (Nezhadali and Zarrabi Shirvan, 2010).

Karyological data are essential information for any organism and many karyological investigations have been performed, providing important characters for plant systematic and evolutionary analyses (Stace, 2000).

Many species of Anthemideae have been studied in karyological comparisons (e.g., Carr et al., 1999; Watanabe, 2002; Valles et al., 2005; Chehregani and Mehanfar, 2008; Chehregani and Hajisadeghian, 2009; Chehregani et al., 2013). Since chromosome number of less than 40% of the species of the Asteraceae has been documented, , more studies are still necessary to improve the knowledge of the family as well as the tribe Anthemideae (Volkova and Boyko, 1986; Valles et al., 2005).

Basic chromosome number within Anthemideae tribe is x=9, though x=8 with different ploidy levels have been reported by some researchers (diploid, tetraploid and hexaploid) (Chehregani et al., 2011; Chehregani and Hajisadeghian, 2009; Chehregani and Mehanfar, 2008; Javadi et al., 2013).

In systematic, chromosome number is an important character for plant evolutionary studies and may provide some information about polyploidy and other highly significant genome changes (Guerra, 2008; Louzada et al., 2010) or the benefits of plant chromosome number databases are useful tools for systematic comparisons of geographical and taxonomical groups of plants (Peruzzi et al., 2012).

Corresponding authors E-mails:
*Javadi@rifr-ac.ir; hjavadim@yahoo.com

Also, chromosome counts can increase our understanding of phylogenetic relationships at different taxonomic levels (Yang et al., 2009).
The purpose of the present study is to determine chromosome number, ploidy levels, karyotype analyses and asymmetry indices in different populations of *Tanacetun polycephalum* Schultz-Bip. and *Tanacetun parthenium* Schultz-Bip. Asymmetry indices are the best knowledge of karyological feature of taxa (Altınordu et al., 2014). In literature, there is no sufficient data about chromosome number and karyotype analyses of studied taxa. We expect that our study will be performed in the light of previous cytotaxonomic studies.

## Materials and Methods

In this study, we used root tip meristems from seedling obtained by the germination of ripe seeds collected from various locations (19 populations) on wet filter paper in petri dishes and have been left at 25°C temperature. The studied populations are listed in table 1.

## Pretreatment and Preparation

Root tip meristems obtained from seedlings were pretreated with 0.05% (w/v) 8-hydroxyquinoline for 4 to 5 h at 16°C. Pretreated root tips were fixed in a 3:1 (v/v) mixture of 95% (v/v) ethanol and propionic acid for 24 h. Root tips were hydrolyzed in 1M HCl for 5 to 7 min at 60°C and stained in Schiff's reagent for 2 h at room temperature.
Feulgen stain was removed and the root tips were rinsed with cold double-distilled water and stained with Carbol fuchsin stain overnight at 4°C in a refrigerator. After staining, the root tips were washed three to four times with cold double distilled water and stored in cold double-distilled water in a refrigerator.
Root tips were squashed in a droplet of 45% (v/v) acetic acid and lactic acid (10:1). The preparations were observed with an optical microscope (BX41 Olympus supplemented with Digital color video camera) at a magnification of about 2000 X. The best plates were selected and captured. For each population, 5 mitotic metaphase plates were prepared.

## Karyotype Analyses

The following parameters were measured in each metaphase plate to characterize the karyotypes numerically: haploid chromosome numbers (n), long arm (LA), short arm (SA), total length (TL=LA+SA), genome size (∑TL), arm ratio (AR=LA/SA), centromeric index [CI=SA/(LA+SA)], difference of range relative length (DRL=MaxRL%-MinRL%), (MaxRL%=[MaxTL/(∑TL)*100], (MinRL%=[MinTL/(∑TL)*100], that MaxRL% and MinRL% are relative length of longer and shorter chromosome respectively, karyotype formula (KF) according to Levan's method (Levan et al., 1964), the classification of chromosomes as median (m), submedian (sm), subterminal (st) and terminal point (T).
For analysis of karyotype asymmetry, the following methods were used. To describe karyotype asymmetry and to determine the relationships of karyotypic between species, Huziwara (1962) developed the total form percent (TF%=[(∑SA/∑TL)*100]). Romero Zarco (1986) also, provided a different method to measure karyotype asymmetry which is the intrachromosomal asymmetry index ($A_1$=1-[∑(SA/LA)/n]), where SA and LA are the mean length of short and long arms of each pair of homologous, respectively and n is the number of homologous.
The other interchromosomal asymmetry index is ($A_2$=s/x), whereas s and x are the average of standard deviation and mean of chromosome length, respectively.
Also, karyotypic evolution has been determined using the symmetry classes of Stebbins (SC) (Stebbins, 1971). Stebbins (1971) distinguished 12 categories concerning the karyotype asymmetry and from which 10 categories were known to occur in higher plants.
He established these by recognizing three degrees of difference (A–C) between the largest and smallest chromosome complement, and four degrees (1–4) with respect to the proportion of chromosomes which are median pair with an arm ratio of less than 2:1.

## Results

The pictures of the mitotic metaphase samples and their karyotypes were presented in Figures 1and 2. The results showed that in all of populations the basic chromosome number was x=9, and showed two ploidy levels (2x and 4x).
The type of all chromosomes usually were metacentric (m), sub-metacentric (sm) and rarely sub-telocentric (st). The somatic chromosome numbers (2n) and karyotypic details for the studied populations were presented in Table 2.

Research Article

**Table 1**. Materials used for chromosomal study of *Tanacetum polycephalum* Schultz-Bip. and *Tanacetum parthenium* Schultz-Bip. populations. (G.B.C:Gene Bank Code, H.C:Hetrbarium Code)

| No. | G.B.C | H.C | Population | Location | altitude (m) | Latitude | Longitude |
|---|---|---|---|---|---|---|---|
| 1 | 1459 | 102810 | *Tanacetum polycephalum* | Hamadan, Malayer | 2250 | 48º 30′ 05″ N | 34º 34′ 41″ E |
| 2 | 17650 | 102820 | *Tanacetum polycephalum* | Qom, Qom | 2310 | 50 08′ 52″ N | 34º 08′ 10″ E |
| 3 | 22613 | 102816 | *Tanacetum polycephalum* | Kordestan, Baneh | 1956 | 45º 57′ 00″ N | 36º 03′ 00″ E |
| 4 | 25991 | 102818 | *Tanacetum polycephalum* | Kordestan, Qorveh | 1995 | 47º 43′ 03″ N | 35º 12′ 46″ E |
| 5 | 31204 | 102814 | *Tanacetum polycephalum* | Kohkiloyeh & Boyerahmad | 2435 | 51º 42′ 37″ N | 30º 03′ 33″ E |
| 6 | 33331 | 102812 | *Tanacetum polycephalum* | Zanjan | 2250 | 48º 13′ 56″ N | 36º 36′ 50″ E |
| 7 | 35122 | 102819 | *Tanacetum polycephalum* | Esfahan, Golpayegan | 1970 | 50º 13′ 39″ N | 33º 28′ 30″ E |
| 8 | 35185 | 102817 | *Tanacetum polycephalum* | West-Azerbaijan, | 1648 | 45º 01′ 55″ N | 38º 01′ 31″ E |
| 9 | 35579 | 102815 | *Tanacetum polycephalum* | Mazandaran, Savadkoh | 2024 | 52º 57′ 16″ N | 36º 51′ 52″ E |
| 10 | 35635 | 102813 | *Tanacetum polycephalum* | Mazandaran, Amol | 2570 | 52º 06′ 00″ N | 35º 52′ 13″ E |
| 11 | 10262 | 102811 | *Tanacetum parthenium* | Yazd, Taft | 2310 | 54º 08′ 01″ N | 31º 37′ 97″ E |
| 12 | 20096 | 102827 | *Tanacetum parthenium* | Qazvin, Qazvin | 1700 | 50º 10′ 00″ N | 36º 26′ 00″ E |
| 13 | 27153 | 102822 | *Tanacetum parthenium* | Gilan, Shaft | 970 | 49º 11′ 12″ N | 36º 56′ 40″ E |
| 14 | 27158 | 102826 | *Tanacetum parthenium* | Gilan, Fuman | 1120 | 49º 02′ 42″ N | 37º 08′ 52″ E |
| 15 | 27162 | 102825 | *Tanacetum parthenium* | Gilan, Astara | 1460 | 48º 33′ 12″ N | 38º 15′ 50″ E |
| 16 | 27173 | 102821 | *Tanacetum parthenium* | Gilan, Talesh | 945 | 48º 42′ 53″ N | 37º 58′ 12″ E |
| 17 | 29813 | 102823 | *Tanacetum parthenium* | Kohkiloyeh & Boyerahmad | 1800 | 51º 10′ 16″ N | 31º 50′ 59″ E |
| 18 | 33183 | 102814 | *Tanacetum parthenium* | Hamadan | 21413 | 48º 29′ 45″ N | 34º 43′ 57″ E |
| 19 | 35190 | 102824 | *Tanacetum parthenium* | West-Azerbaijan, | 2167 | 46º 41′ 72″ N | 36º 51′ 36″ E |

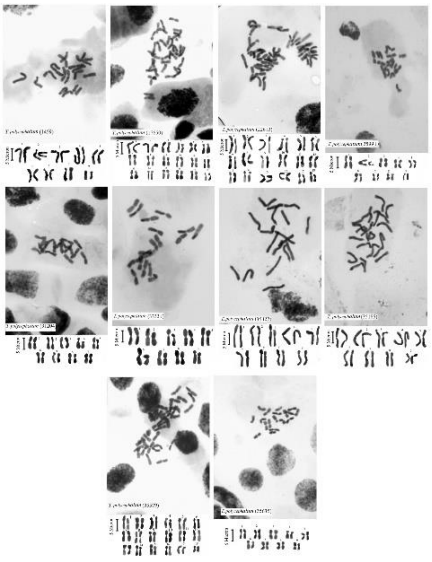


**Figure 1**. The mitotic metaphase samples and their karyotypes

**Table 2.** Karyotype characteristics of *Tanacetum polycephalum* Schultz-Bip. and *T. parthenium* Schultz-Bip. populations. 2n- somatic chromosome number, SC- symmetry classes of Stebbins, ∑TL-length of genome(micron), TL-mean of chromosome Length (micron), AR- arm ratio, CI- Centromer Index, TF%- total form percentage, $A_1$-intrachromosome asymmetry index, $A_2$- interchromosome asymmetry index, DRL- difference of relative length, KF-karyotype Formula (m: metacentric, sm: submetacentric, st: subtelocentric)

| N | Populat | 2 | S | ∑T | TL | A | CI | TF | $A_1$ | $A_2$ | DR | KF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *Tanacet* | 1 | 2 | 73.4 | 8.1 | 2.0 | 0.3 | 37. | 0.3 | 0.2 | 7.9 | 5m(1,2,4,5,6)+3Sm(3,8,9)+1St(7) |
| 2 | *Tanacet* | 3 | 2 | 100. | 5.5 | 1.7 | 0.3 | 37. | 0.3 | 0.1 | 3.9 | 10m(3,5,7,8,9,10,11,12,13,16)+7Sm(2 |
| 3 | *Tanacet* | 3 | 2 | 129. | 7.1 | 1.7 | 0.3 | 37. | 0.3 | 0.1 | 3.3 | 11m(2,5,6,8,10,11,12,13,15,16,17)+5S |
| 4 | *Tanacet* | 1 | 2 | 46.7 | 5.1 | 1.5 | 0.3 | 39. | 0.3 | 0.2 | 7.2 | 7m(1,2,3,4,5,6,8)+2Sm(7,9) |
| 5 | *Tanacet* | 1 | 2 | 54.9 | 6.1 | 1.4 | 0.4 | 42. | 0.2 | 0.0 | 2.3 | 7m(1,2,3,5,6,8,9)+2Sm(4,7) |
| 6 | *Tanacet* | 1 | 2 | 81.3 | 9.0 | 1.4 | 0.4 | 42. | 0.2 | 0.1 | 3.9 | 7m(1,2,4,5,6,7,8)+2Sm(3,9) |
| 7 | *Tanacet* | 1 | 2 | 94.3 | 10. | 2.0 | 0.3 | 35. | 0.4 | 0.1 | 5.0 | 4m(1,2,4,5)+4Sm(3,6,7,9)+1St(8) |
| 8 | *Tanacet* | 1 | 2 | 95.4 | 10. | 1.6 | 0.3 | 39. | 0.3 | 0.2 | 8.3 | 6m(1,2,3,5,7,9)+2Sm(4,8)+1St(6) |
| 9 | *Tanacet* | 3 | 2 | 110. | 6.1 | 1.6 | 0.3 | 39. | 0.3 | 0.1 | 4.2 | 14m(1,2,5,6,7,8,10,11,12,13,14,16,17, |
| 1 | *Tanacet* | 1 | 2 | 45.6 | 5.0 | 1.3 | 0.4 | 43. | 0.2 | 0.1 | 3.2 | 8m(1,2,3,5,6,7,8,9)+1Sm(4) |
| 1 | *Tanacet* | 1 | 2 | 59.1 | 6.5 | 1.7 | 0.3 | 38. | 0.3 | 0.3 | 13. | 5m(1,4,7,8,9)+3Sm(2,3,5)+1St(6) |
| 1 | *Tanacet* | 1 | 2 | 50.5 | 5.6 | 1.7 | 0.3 | 38. | 0.3 | 0.1 | 4.9 | 4m(1,2,3,7)+5Sm(4,5,6,8,9) |
| 1 | *Tanacet* | 1 | 2 | 44.0 | 4.8 | 1.5 | 0.4 | 40. | 0.3 | 0.1 | 4.5 | 6m(1,2,3,4,5,6)+3Sm(7,8,9) |
| 1 | *Tanacet* | 1 | 2 | 57.4 | 6.3 | 1.6 | 0.3 | 38. | 0.3 | 0.1 | 4.7 | 5m(3,4,5,6,9)+4Sm(1,2,7,8) |
| 1 | *Tanacet* | 1 | 2 | 44.2 | 4.9 | 1.5 | 0.3 | 39. | 0.3 | 0.1 | 6.1 | 7m(1,3,5,6,7,8,9)+2Sm(2,4) |
| 1 | *Tanacet* | 1 | 2 | 39.0 | 4.3 | 1.6 | 0.3 | 38. | 0.3 | 0.2 | 6.3 | 6m(1,2,4,6,7,9)+3Sm(3,5,8) |
| 1 | *Tanacet* | 3 | 2 | 110. | 6.1 | 1.6 | 0.3 | 38. | 0.3 | 0.2 | 4.8 | 11m(2,5,6,8,9,10,11,14,16,17,18)+6S |
| 1 | *Tanacet* | 1 | 2 | 43.2 | 4.8 | 1.6 | 0.3 | 39. | 0.3 | 0.2 | 8.1 | 6m(3,4,5,6,7,8)+3Sm(1,2,9) |
| 1 | *Tanacet* | 3 | 2 | 74.2 | 4.1 | 1.3 | 0.4 | 42. | 0.2 | 0.1 | 3.9 | 17m(1,2,3,4,5,6,7,8,10,11,12,13,14,15, |

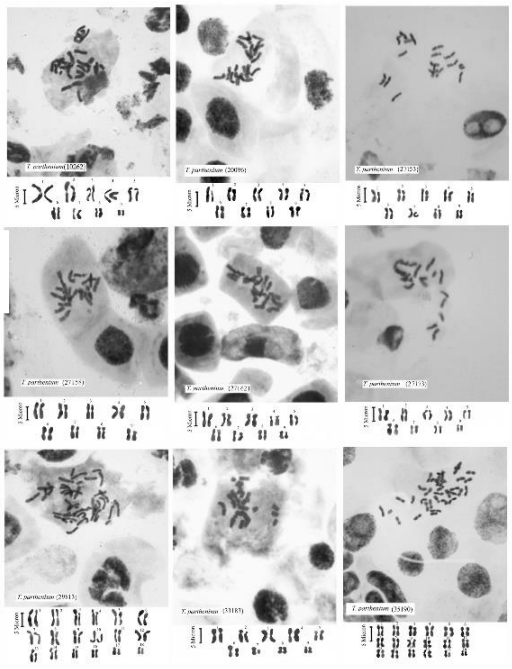


**Figure 2.** The mitotic metaphase samples and their karyotypes

In the present work, we have provided additional karyomorphological parameters using Symmetry of Stebbin's (SC), Total Form percentage (TF%), asymmetry indices of Romro-Zarco ($A_1$ and $A_2$) (Romero, 1986) and Difference Relative Length (DRL), which do not depend on chromosome number or chromosome size. The scatter diagram of populations dispersion based on two components ($A_1$–$A_2$) with Stebbins symmetry (SC) are represented graphically in Figure 3.

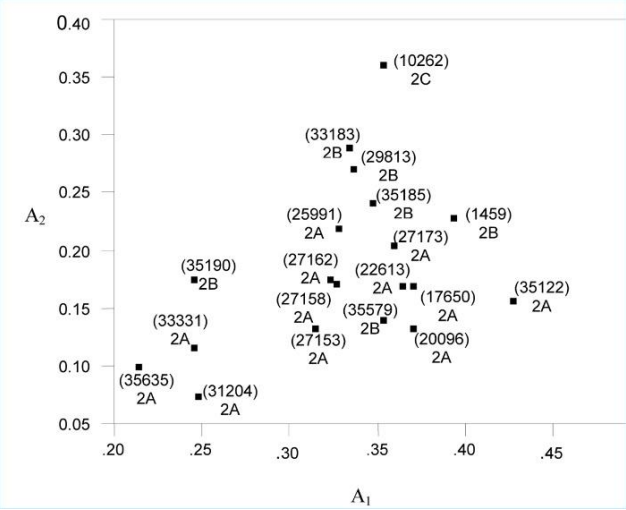



**Figure 3.** The scatter diagram of populations dispersion based on two components ($A_1$–$A_2$) with Stebbins symmetry (SC).

All populations are located in Stebbin's classes (SC) 2A and 2B, except for *T. parthenium* (10262), with the highest DRL value (13.615), and $A_2$ (0.363), is located in 2C Stebbin's classes. *T. polycephalum* (35122) and *T. polycephalum* (1459) had the highest value of AR and $A_1$, respectively, while *T. polycephalum* (35635) had the lowest AR value (1.343), and $A_1$ value (0.214). The differences in AR and $A_1$ values among the other populations were not significant (AR=1.355-1.799, $A_1$=0.245-0.371). While the highest TF% value (43.683) and CI value (0.435) belonged to *T. polycephalum* (35635), *T. polycephalum* (35185) and *T. polycephalum* (35122) had the highest of chromosome length (TL) in all populations. So these two populations have the longer chromosome than other populations.

The current available chromosome data showed polyploidy to be the most significant evolutionary trend in chromosome number within the Asteraceae (Carr et al., 1999; Valles et al., 2005; Chehregani et al., 2013).

In this survey, we study variation of chromosome number in *Tanacetum polycephalum* Schultz-Bip. and *Tanacetum parthenium* Schultz-Bip. populations from natural gene bank. We had 10 populations from *T. polycephalum* and nine populations from *T. parthenium*. In all populations the basic chromosome number was x=9 and the populations showed two ploidy levels: Diploid and Tetraploid (Table 2 and Figures 1 and 2). This is the first report of chromosome number in *Tanacetum polycephalum* and *T. parthenium* for populations in natural resources gene bank of Iran, indicating diploid and tetraploid level (2x and 4x) based on x=9. Our results agree with other results, showing the basic chromosome number is 9 with different ploidy levels in Anthemidaea tribe (Goldblat and Johnson, 1979; Torrel et al., 2001; Valles et al., 2005; Yousefzadeh et al., 2010).

The diploid chromosome number (2n=2x=18) has already been reported in *T. albipannosum* Hub.-Mor and Grierson, *T. macrophyllum* Sch. Bip., *T. coccineum* (Willd.) Grierson ssp. chamaemelifolium (Sommier and Levier) Grierson, and *T. sorbifolium* (Boiss.) Grierson from Turkey (Inceer and Hayirlioglu-Ayaz, 2007).

Sonboli and et al., (2011) showed that in *Tanacetum fisherae*, from Kerman province, the constant chromosome number found in all metaphase plates was 2n=44+1B that indicates a pentaploid level (5x) based on x=9. A new ploidy level (pentaploidy) is reported for the first time for the genus.

Chehregani and et al., (2011) showed that in 14 populations of *Tanacetum polycephalum* located in west region, the number of chromosomes was 18 (Diploid), 36 (Tetraploid) and 54 (Hexaploid). Some populations were mixoploidy, such as Diploid and Tetraploid or Tetraploid and Hexaploid.

Finally, in Anthemideae tribe both x=9and x=8 were reported. It seems polyploidy is the cause of polymorphism in this species (Valles et al., 2005; Chehregani and Hajisadegian, 2009).

In Table 2 we showed the Length of Genome (∑TL) existed in tetraploid populations including *T. polycephalum* (Qom, Qom), *T. polycephalum* (Kordestan, Baneh), *T. polycephalum* (Mazandaran, Savadkoh), *T. parthenium* (Kohkiloyeh and Boyerahmad) and *T. parthenium* (West-Azerbaijan, Shahindej) The values of ∑TL are high and the highest value of ∑TL belonged to *T. polycephalum* (Kordestan, Baneh). In diploid populations, the highest value of ∑TL is shown in *T. polycephalum* (West-Azerbaijan, Uromeyeh) and *T. polycephalum* (Esfahan, Golpayegan). So, in tetraploid populations, *T. polycephalum* (Kordestan, Baneh) have the longest chromosome (TL=7.197µm) and in diploid populations, *T. polycephalum* (West Azerbaijan, Uromeyeh) and *T. polycephalum* (Esfahan, Golpayegan), have the longest chromosome with TL=10.602µm and

TL=10.482µm, respectively. The smallest chromosome length (4.124µm) is observed in the population of *T. parthenium* (West-Azerbaijan, Shahindej). Therefore the population of *T. polycephalum* has the longest chromosomes (bigger) than the population of *T. parthenium*.

When we compare the karyotype asymmetry in genus *Tanacetum* according to Stebbins (1971) classification, all of the populations are located in Stebbin's classes (SC) 2A and 2B, except for *T. parthenium* (Yazd, Taft), that is classified to symmetry classes of Stebbins as 2C. Among these classes, 2A class is more symmetrical than 2B class and 2B class is more symmetrical than 2C class, but we cannot determine which 2A class has higher symmetry. So Stebbins classification does not clarify this situation. In order to determine the most symmetrical or asymmetrical karyotype we used other indices. The karyotype asymmetry was evaluated based on six different parameters includes AR (arm ratio), CI (centromeric index), TF% (total form percentage), $A_1$ and $A_2$ (intrachromosome and interchromosome asymmetry index), DRL (difference relative length). CI and TF% have a direct relation with symmetry of karyotype based on type of chromosomes. Increase of the CI and TF% values is associated with increased symmetry in karyotype and vice versa. . The AR, $A_1$–$A_2$ and DRL values increase with increasing of asymmetry (Zuo and Yuan, 2011).

The AR and $A_1$ are the asymmetric parameters showing the chromosomes are not the same in type. So according to AR and $A_1$ indices, *T. polycephalum* (Mazandaran, Amol) has the most symmetrical karyotype and *T. polycephalum* (Esfahan, Golpayegan) has the most asymmetrical karyotype. Also, the karyotype formula confirmed this issue; in population of *T. polycephalum* (Mazandaran, Amol) karyotype formula is 8m+1sm, in which 8 chromosomes are metacentric and one chromosome is sub-metacentric.

*T. parthenium* (Yazd, Taft) had the highest of $A_2$ and DRL values. $A_2$ and DRL also are asymmetric parameter and show that karyotype is asymmetric based on length of chromosomes. So population of *T. parthenium* (Yazd, Taft) have an asymmetric karyotype based on the length of chromosomes and its classification in 2C Stebbins classes.

In addition to various symmetrical states by Stebbins, the variance of different populations according to $A_1$ and $A_2$ values are presented in Figure 3. With regard to Figure 3, the pattern of variation of $A_1$ and $A_2$ values have been compared with the pattern of Stebbins system. This diagram also shows that the *T. parthenium* (Yazd, Taft) has the most derived karyotype.

In the present research, somatic chromosome number, karyotype analysis and karyotype asymmetries of 19 popilations *T. polycephalum* and *T. parthenium* from the family of Asteraceae were defined for the first time. This study will play a positive role to enlighten this taxonomically revision genus (Taner et al., 2014). Determination of the number of chromosomes in the genus *Tanacetum* also shed light on the opinions of further studies in this regard.

## Discussion and Conclusions

In all of the populations the basic chromosome number was x=9, and showed two ploidy levels (2x and 4x).

The populations of *Tanacetum polycephalum* have the longest chromosomes than populations of *T. parthenium*.

The asymmetric and symmetric karyotype is shown in populations of *T. polycephalum* from Esfahan-Golpayegan and *T. polycephalum* from Mazandaran-Amol, respectively.

Populations of *T. parthenium* from Yazd-Taft with the longer and smaller chromosome (highest of $A_2$ and DRLvalue) also have an asymmetric karyotype.

## Acknowledgment

The authors are grateful to the Gene Bank for providing of seeds and the Research Institute of Forests and Rangelands (RIFR) in Iran for financial support.

## References

1. Altınordu F. and Martin E. (2014) New chromosome counts, karyotype analyses and asymmetry indices in some taxa of genus *Senecio* L. and related genera *Tephroseris* (Rchb.) Rchb. and *Turanecio* Hamzaoglu belong to tribe Senecioneae (Asteraceae) from Turkey. Plant Systematics and Evolution 300: 2205–2216.
2. Carr GD., King RM., Powell AM. and Robinson H. (1999) Chromosome number in Compositeae XVIII. American Journal of Botany 86: 1003-1013.
3. Chehregani A., Atri M., Sarmadi J. and Asgari M. (2011) Chromosome number variation in *Tanacetum polycephalum* Schultz bip. (l.) (Asteraceae) in west of Iran. Caryologia 64(3): 302-308.
4. Chehregani A. and Hajisadeghian S. (2009) New chromosome counts in some species of Asteraceae from Iran. Nordic Journal of Botany 27: 247-250.
5. Chehregani A., Heidarian M. and Jalali F.

(2013) Cytological Investigations on *Achillea wilhelmsii* C. Koch (Asteraceae) in the Western part of Iran. Chromosome Botany 8:13-16.

6. Chehregani A. and Mehanfar N. (2008) New chromosome counts in the tribe Anthemideae (Asteraceae) from Iran). Cytologia 73(2):189–196.
7. Goldblatt P. and Johnson DE. (1979) The family of Asteraceae. Index to plant chromosome numbers. Missouri Botanical Garden, St. Louis. Available in www.tropicos.org.
8. Guerra M. (2008) Chromosome numbers in plant cytotaxonomy: concepts and implications. Cytogenetic and Genome Research 120: 339–350.
9. Heywood VH. and Humphries CJ. (1977) Anthemideae-systematic review. In: Heywood,V.H.R., Harborne, J.B. and Turner, B.L. (eds). The biology and chemistry of the Compositae. Volume II. Academic Press, Great Britain. pp. 851-897.
10. Hulten E. (1968) Flora of Alaska and Neighboring territories. A manual of the vascular plants. Stanford University Press, Stanford, California. 1008 pp.
11. Huziwara Y. (1962) Karyotype analysis in some genera of compositae. VIII Further studies on the chromosomes of Aster. American Journal of Botany 49:116-119.
12. Inceer H. and Hayirlioglu-Ayaz S. (2007) Chromosome numbers in the tribe Anthemideae (Asteraceae) from north-east Anatolia. Botanical Journal of Linnaean Society 153(2): 203–211.
13. Javadi H., Alizadeh MA., Emami H. and Salehi Shanjani P. (2013) Cytogenetics investigation in different populations of *Anthemis tinctoria* L. (yellow chamomile) accessions from Iran. Chromosome Science 16: 43-49.
14. Levan A., Fredge K. and Sanberge A.A. (1964) Nomenclature for centromeric position on chromosomes. Hereditas 52: 201-220.
15. Louzada RB., Palma-Silva C., Corrêa AM., Kaltchuk-Santos E. and Wanderley MDGL. (2010) Chromosome number of *Orthophytum* species (Bromeliaceae). Kew Bull 65: 53-58.
16. Mitich LW. (1992) Intriguing world of weeds-Tansy. Weed Technology 6: 242-244.
17. Mozaffarian V. (2006) A Dictionary of Iranian Plant Names. Latin-English-Persian. Fahang Moaser, 671 pp.
18. Nezhadali A. and Zarrabi-Shirvan B. (2010) Separation, identification and determination of volatile compounds of *Ziziphora persica* bunge using HS-SPME/GCMS. International Journal of Environmental Science and Development 1(2): 115–118.
19. Peruzzi L., Góralski G., Joachimiak AJ. and Bedini, G. (2012) Does actually mean chromosome number increase with latitude in vascular plants? An answer from the comparison of Italian, Slovak and Polish floras. Comparative Cytogenetics 6: 371–377.
20. Romero Zarco C. (1986) A new method for estimating karyotype asymmetry. Taxon 36: 526-530.
21. Sonboli A., Olanj N. and Pourmirzaei A. (2011) Biosystematics and phylogeny of *Tanacetum fisherae*, a new record from Iran. Rostaniha (Botanical Journal of Iran) 12 (2): 165-175.
22. Soreng R. J. and Cope E. A. (1991) On the taxonomy of cultivated species of the Chrysanthemum genus-complex (Anthemideae: Compositae). Baileya 23:145-165.
23. Stace CA. (2000) Cytology and cytogenetics as a fundamental resource for 20th and 21th centuries. Taxon 49: 451-477.
24. Stebbins GL. (1971) Chromosomal evolution in higher plants. Edward Arnold Publisher, London, Ltd. p.216.
25. Taner Ö., Gezer E. and Martin E. (2014) Karyotype analyses on the genus *Lallemantia* fisch. & C. A. Mey. (Lamiaceae) from Turkey. Cytologia 79(4): 553–559.
26. Torrel M., Valles J., Garcia-Jakas N., Mozaffarian V. and Gab-rielian E. (2001) New or rare chromosome counts in the genus *Artemisia* L. (Asteraceae, Anthemideae) from Armenia and Iran. Botanical Journal of the Linnean Society 135: 51–60.
27. Tutin T., Heywood V., Burges N., Moore D., Valentine D., Walters S. and Webb D. (1976) Flora europaea: plantaginaceae to compositae (and Rubiaceae). Vol. 4. Cambridge University Press, Cambridge.
28. Valles J., Garnatje T., Garcia S., Sanz M. and Ko-Robkov A. A. (2005) Chromosome numbers in the tribes Anthemideae and Inuleae (Asteraceae) in Kazakhstan. Botanical Journal of Linean Society 148: 77-85.
29. Volkova S. A. and oyko E. V. (1986) Chromosome numbers in some species of Asteraceae from the southern part of the Soviet Far East. Botanicheskii Zhurnal 71: 16-93.
30. Watanabe K. (2002) Index to chromosome numbers in Asteraceae. Available in http://www-asteraceae.cla.kobe-u.ac.jp/ index.html.
31. Watson L. E. (1754) Flora of North America, *Tanacetum*. Vol, 19, 20 and 21. Available in www.eFloras.org.
32. Yang Z., Zhang L., Zhao H., Yang R., Ding C., Zhou Y. and Wan D. (2009) Chromosome numbers of some species of *Salvia*

(Lamiaceae) from the Sichuan Province, China. Nordic Journal of Botany 27: 287–291.

33. Yousefzadeh K. Hoshmand S. and Zinali H. (2010) Karyological study four speices of *Anthemis* from Iran. Iranian Journal of Rangelands and Forests Plant Breeding and Genetic Research 18(1): 55-62.
34. Zuo L. and Yuan Q. (2011) The difference between the heterogeneity of the centromeric index and intrachromosomal asymmetry. Plant Systematics and Evolution 297: 141–145.

*Journal of Cell and Molecular Research (2017) 9 (2), 67-72* DOI: 10.22067/jcmr.v9i2.69816

Research Article

# The Effects of Silveroxide on mRNA Level of Key Genes Involved In the Biosynthesis of RebaudiosideA in *Stevia Rebaudiana*

Alireza Mirzaei[1], Sara Dezhsetan[1], Gholamreza Kavoosi[*2], Mahdi Behnamian[3]

[1] *Department of Agronomy and Plant Breeding, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran*

[2] *Department of Biotechnology, College of Agriculture, Shiraz University, Shiraz, Iran*

[3] *Department of Horticultural Sciences, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran*

Received 1 December 2017 Accepted 20 December 2017

## Abstract

*Stevia rebaudiana* bertoni is a plant with sweetening properties. This medicinal plant is perennial and from *Asteraceae* family. Its leaves contain glycoside compounds of a sugar part and non-sugar sectors. One of the glycosides compounds is RebaudiosideA which has a greater importance in market. Several key regulating genes including copalyl diphosphate synthase (*CPPS*) (AF034545.1), geranylgeranyl diphosphate synthase (*GGDPS*) (DQ432013.3), (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase (*HDS*) (FJ755689.1), UDP glucosyltranserase 85C2(*UGT85C2*)(AY345978.1), UDPglucosyltranserase–74G1 (UGT74G1) (AY345982.1) and UDP glucosyltranserase–76G1 (UGT76G1) (KC631816.1) are involved in the biosynthesis Rebaudioside A. This experiment was conducted to evaluate the effect of silver oxide $Ag_2O$ on the mRNA level of these genes in the *stevia rebaudiana*. The experiments repeated 3 times and with concentrations of 50, 100 and 200 µM. Increasing concentrations of 50 micromoles of silver oxide up to 100 micromoles leads to an increase in the expression levels of all the studied genes. Also according to the expression profile of these genes and the results of HPLC there is a significant increase on the expression level of the genes and production of RebaudiosideA under $Ag_2O$ treatment. In general, it was found that increasing the concentration of $Ag_2O$ can lead to an increase in the level of mRNA for the chosen genes. On the other hand, the low expression of the genes studied under control conditions (No Treatment), compared to the treatment with $Ag_2O$, revealed that the treatment can lead to higher sweetener glycoside components in the Stevia leaves. The physiological assay showed that $Ag_2O$ treatment in concentrations of 100 and 200 µM have more positive effect on chlorophyll, protein, carbohydrates and carotenoids in Stevia.

***Keywords*:** Stevia, RebaudiosideA, Key genes, Gene expression

## Introduction

In recent years, the development of natural sweeteners with characteristics of: low calorific, non-cariogenic and healthier is on rise. One of the promising alternatives is diterpenoid glycosides family, especially steviol glycoside. They are usually found in leaves of *Stevia rebaudiana* (Soejarto, 2002). Effects of its therapeutic properties comes from the presence of phenolic compounds in different parts of this plant, especially in leaves and callus, and it has anti-cancer effects and possible treatment of cardiovascular and blood sugar diseases by its property absorption is considered by free radicals (Tadhani et al., 2007). In 1931, two french chemists named Bridel and Lavieille succeeded in extracting the sweetener of the plant called steviol glycoside. They obtained natural steviol glycoside by a series of non-enzymatic hydrolyzing experiments (Bridel and Lavieille, 1931).

In fact, steviol glycoside is a diterpene glycoside that has been identified in the extract of various constituents of stevia rebaudiana and is the most important sweetener in stevia (Dacomi et al., 2005). Steviol glycoside contains compounds such as DalcosideA, RibaudiosideE-A, Steviolbioside, Stevioside, etc. Currently, 34 compounds are known in steviol glycoside, from which, 8 of them are in the form of isomers. Among these compounds, RibaudiosideA and Stevioside are more important than other compounds and are recognized as dominant glycoside compounds in Stevia rebaudiana leaf extract and they have plenty of commercial significance (Chaturvedula et al., 2011). Stevia also contains protein, fiber, carbohydrates, phosphorus, iron, calcium, sodium, potassium, magnesium, zinc and vitamins A and C. These sweet compounds pass through the digestive processes without chemical

Corresponding authors E-mail:
* ghkavoosi@shirazu.ac.ir

decomposition and are therefore useful for those who need to control their blood glucose (Rita, 1997). Since metal oxides are capable of stimulating the mechanism of defense in the cell and the production of secondary metabolites, they have recently been used in culture media as Elicitors. Lobick (2008) has shown that silver oxide, through the production of active oxygen species (ROS), affects the membrane phospholipids of the microorganisms, causing their peroxidation and degradation of the cellular membrane of the microorganisms. Also silver oxide with sulfur group in the membrane of the microorganisms is replaced and creates pores in the cellular membrane of the microorganisms that cause leakage of the intracellular material of the microorganisms and their destruction (Dibrov et al., 2002). Today, metal oxides serve as an attractive candidate for the delivery of many small molecules or large biochemical molecules (Jiang et al., 2006). Among various metal oxides, silver oxide is being used for its good conductivity (Jiang et al., 2006), chemical stability (Mafune et al., 2000), catalytic activity (Jain et al., 2008) and antimicrobial property (Jeeva et al., 2014). Indeed, the discovery of these basic properties of silver oxide has led to its importance in biology and biotechnology (Jia et al., 2013). Accordingly, this study aimed to investigate the effect of silver oxide, as an elicitor, on the mRNA level of the (AF034545.1) *CPPS*, (FJ755689.1) *HDS*, (DQ432013.3) *GGDPS*, (AY345978.1) *UGT85C2*, (AY345982.1) *UGT74G1* and (KC631816.1) *UGT76G1* genes in Stevia.

## Materials and Methods

### Plant Growth

Very small and thin (approximately 3 mm) light brown seeds of stevia (*Stevia rebaudiana* bertoni) were purchased from Techno Kesht Shiraz Company (Shiraz, Iran). These seeds were surface-sterilized with sodium hypochlorite (5%). Then, a piece of filter paper was put into each petri dish and 5 ml of distilled water was added. The seeds were transferred onto the filter paper, with 10 seeds in each of the 32 petri dishes. About 1cm space was kept between the seeds. The dishes were covered with their lids and sealed with parafilm. After preparing the media and sealing the dishes, all petri dishes were placed in an incubator. More than 80% of the seeds germinated in the dark within five days at room temperature. The seedlings developed roots with at least 25 mm length and were then transferred into the jardiniere containing cocopeat and perlite with proportion 70 to 30, respectively. During the experimental period, the jardiniers were irrigated twice a week. The seedlings received tap water at greenhouse conditions with about 60% relative humidity for two weeks until the size of leaves reached about 10cm. Diurnal cycles at each temperature (27/24°C) were set at 16h light and 8 h dark cycles. At this state, the seedlings were treated with 50 mL of different concentrations (50, 100, 200 µM) of silveroxide, loaded onto the gelatin dispersion. The equal concentrations of gelatin dispersion were used as control conditions which continued for three days. Young leaves were harvested and frozen in liquid nitrogen for twenty-four hours after each treatment, and then stored at -70°C for the RNA extraction and real-time PCR analysis. For biochemical analysis, the young leaves were harvested and stored at 4°C.

### Primer Design

Primers were designed using AlleleID7 software (Premier Bio soft Intl, Palo Alto, CA, USA) for the target genes as shown in Table 1. Elongation factor genes were used as the internal control (whose expression proved not to be influenced by metal oxides treatment) for data normalization (9, 10). Two primer pairs were designed for each individual gene, as represented in Table 1. Primers for the PCR reactions were designed to have a melting temperature of about 50°C to 55°C and to give a PCR product between 100 and 200 bp in length.

**Table 1**. Sequences of primers used for Real-Time PCR amplification and the expected product size.

| Primer | Sequence | length of the piece on cDNA | TM |
|---|---|---|---|
| UGT85C2F | ATGATGTATTGGACACTTGCTG | 166 | 54.2 |
| UGT85C2R | CCTTGAGACGGATGCCTTC | 166 | 54.7 |
| UGT76G1F | CATCTTTCACACCAACTTCAAC | 178 | 53.5 |
| UGT76G1R | GCGTCGTAATTCGTCAGC | 178 | 53.5 |
| UGT74G1F | CCTTGGTCTCAGATGTTGTTTG | 105 | 55 |
| UGT74G1R | CGTCCACTCTATTACCTCTTCC | 105 | 55 |
| CPPSF | CTACACGGCTTCGCTTTG | 113 | 53.1 |
| CPPSR | GTCACATCTACTCCATCTTGC | 113 | 53.4 |
| GGDPSF | CGATTGGTTTGTTGTTTCAG | 169 | 50.8 |
| GGDPSR | GCTTCCTTGTTTAATTTCTCC | 169 | 50.4 |
| HDSF | TTTCTTGGCTCCGTATCG | 178 | 50.9 |
| HDSR | TGAGGCTACATCTGAATAGG | 178 | 50.6 |
| Elongation Factor-F | GATGCTTCCGACTAAACCTATGG | 113 | 50.6 |
| Elongation Factor-R | CACTCTTGATAACACCGACTGC | 113 | 50.9 |

### RNA Preparation and cDNA Synthesis

Total RNA was extracted from 100 mg of leaf material, using RNX-Plus buffer (Cinnagen, Tehran, Iran) according to the manufacturer's instructions. The quantification of the total RNA was performed with a Nano-drop (ND 1000 spectrophotometer) at 260 nm (Thermo Fisher scientific, wilmington, DE, USA). The RNA integrity was checked by visual observation of 28S rRNA and 18S rRNA bands on

an agarose gel electrophoresis before Real-time PCR analysis (Figure 1). In order to ensure the absence of the DNA contamination, DNase treatment (Fermentas Company) was performed on the extracted RNAs. Then synthesis of the first cDNA strand synthesis was performed, using commercial kit (Fermentas company).

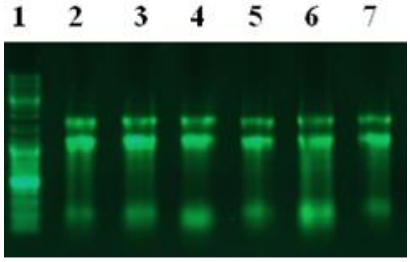



**Figure 1**. Agrosoe gel electrophoresis of the total RNA extraction from the Stevia leaves. 1. Gene RulerTM DNA Ladder Mix (Fermentas); RNAs extracted from leaves stevia for genes; 2. *HDS*: (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase; 3. *GGDPS*: geranylgeranyl diphosphate synthase; 4. *CPPS*: copalyl diphosphate synthase; 5. *UGT85C2*: UDP glucosyltranserase – 85C2; 6. *UGT74G1*: UDP glucosyltranserase – 74G1; and 7. *UGT76G1*: UDP glucosyltranserase – 76G1.

**Biochemical Analysis**

Chlorophyll a, b, Total chlorophyll and carotenoid content of the leaves were determined according to Arnon (1967) method. Pigments were extracted in 80% cold acetone and the absorbance of the extractions was measured spectrophotometrically at 645, 663, and 470 nm wavelength and subsequently, pigmental content were determined, based on the following standard formulas (Kichtenthaler and Wellburn, 1983) :

(1) Total chlorophyll (mg/ml) = 20.2 (A645) + 8.02 (A663)
(2) Chlorophyll a (mg/ml) = 12.7 (A663) - 2.69 (A645)
(3) Chlorophyll b (mg/ml) = 22.9 (A645) - 4.68 (A663)
(4) Carotenoid (mg/ml) = (1000A470 - 3.27[Chl a] - 104[Chl b])/227

The soluble carbohydrates were measured in a similar way of proline measurement after extraction. 0.1 ml Alcoholic extract with 3 ml anthrone were freshly prepared (150 mg anthrone+ 100 ml sulfuric acid 72%) and mixed. This solution was placed in water bath for ten minutes and the absorption rate was measured with a spectrophotometer in the wavelength of 625 nm. The amount of the soluble sugars were then calculated (Pereira et al., 1993). For protein extraction, 0.5 g of the dry explant weighed and 4 ml of buffer Tris-HCl was added. Then the samples were thoroughly vortex-mixed shaker for 20 minutes and centrifuged at 5000 rpm for 30 minutes and the upper phase, containing protein, was isolated. For measuring protein content the Bradford method was used. 0.1 ml protein extract of each sample was vortex-mixed with 5 ml of the Bradford solution and its absorbance at 595 nm was recorded (Bradford, 1976).

**Real-time Quantitative PCR Analysis**

Real-time PCR was performed using a line Gene K Thermal cycler (Bioer Technology Co, Hangzhou, China). The cDNA samples were diluted 1:5 by using nuclease-free water, and 5 µL of the cDNA was used for real-time PCR. The final volume for relative real-time PCR was 20 µL containing 4 pmol of each primer, 5 µL (diluted) of the first-strand cDNA and 1x SYBR Premix Ex Taq TM Π (Takara, Japan). The initial denaturing time was 5 min., followed by 40 PCR cycles consisting of 94°C for 10 s, annealing temperatures of each primer 15 s, and 72°C for 30 s. A melting curve was run after the PCR cycles followed by heating from 50 to 95°C. A proper control reaction was carried out without the reverse transcriptase treatment. For each sample, the subsequent real-time PCR reactions were performed in twice under identical conditions (Livak and Schmittgen, 2001).

**Data Normalization and Quantitative PCR Verification**

For real-time data analysis, the relative expression of the target gene in each sample was compared with the control sample (corresponding to the control plants) and was determined with the delta-delta Ct method (Livak and Schmittgen, 2001) with following equations:

$2^{-[\Delta Ct\ sample} - \Delta^{Ct\ control}$ , $\Delta Ct = CT_{target\ gene} - CT_{internal\ control}$

The Ct for each sample was calculated using the Line-gene K software (fqdpcr ver. 4.2.00) and the Larionov (2005) method, where refers to the threshold cycle determined for each gene in the exponential phase of PCR amplification. In this analysis method, the relative expression of the target gene in the control sample was equal to one ($2^0$) by definition.

**HPLC Analysis**

Standard solution of RebaudiosideA was prepared in a mix of 30% water and 70% acetonitrile. To prepare calibration curve, a solution containing 100 µg/mL of the standard in the ratio of 30% water and 70% acetonitrile was used. In order to extract Rebaudioside-A, the *Stevia rebaudiana* leaves were pulverized and nearly 0.1 g of the powder was put in a 20-ml glass vial.

Ten ml of 30% water and 70% acetonitrile mix was added to the vial which was then vortexed. Subsequently, this crude extract was sonicated for 5 minutes. The contents of the vial were centrifuged and the supernatant was diluted 10 times with 30% water

and 70% acetonitrile mix. Finally, 10 microliters of the sample solution were injected into the column for analysis, using KNAUER – smart line manager-5000 HPLC machine. The flow rate was 1 mL per minute. HPLC column profile contains eurospher 100-5 C18 column, 2500-KNAUER Detector, column length 250 mm and measure pore 5µm. During the 14-minute period, the peak of RebaudiosideA appeared at the minute of 2.3 and the wavelength was 220 nm (Kailasam, 2011).

## Results and Discussion

In this research, the effects of silver oxide on the expression pattern of the 6 key regulatory genes involved in biosynthesis of RiboidiosideA in *Stevia rebaudiana* were assayed by Real-time PCR reaction in 3 replicates with concentrations of 50, 100 and 200 µm.

### mRNA Level of the *CPPS*, *GGDPS* and *HDS* Genes

The Duncan's comparison test (at 0.05%) showed that silver oxide treatments at concentrations of 100 µm caused significant changes in the mRNA level of the ***CPPS*, *GGDPS* and *HDS*** genes (Table 2).
Regarding these results, it will be determined that silver oxide treatment will have a positive effect on the increase of gene expression and production of RiboidiosideA, and the best treatment for this gene group will be at a concentration of 100 µm.

**Table 2**. The results of means comparison by Duncan at P<0.05 employing SPSS software version 16

| GENES / Concentrations | CPPS | HDS | GGDPS | UGT76G1 | UGT74G1 | UGT85C2 |
|---|---|---|---|---|---|---|
| 50 µM | 2.9835 B | 3.1956 B | 1.1573 B | 3.1382 A | 2.7647 A | 1.9647 A |
| 100 µM | 8.6395 A | 7.2388 A | 3.0943 A | 3.2156 A | 2.9773 A | 1.6463 A |
| 200 µM | 1.3942 B | 1.1469 B | 1.4601 B | 3.3946 A | 3.2573 A | 2.2862 A |

### Expression of UGT Family Genes

As shown in Figure 2, the amount of genes expression in the treatment of silver oxide compared to the control conditions or non-treatment conditions has increased significantly.
Also increasing the concentration of silver oxide treatment leads to an increase in the expression of UGT family genes.
Generally, the highest amount of genes expression related to concentration of 200 µm. In the mode of comparison between the genes of this family, the *UGT76G1* gene showed the highest expression levels, and the *UGT85C2* gene had the lowest expression.
Regarding the changes in the expression of UGT family genes, it can be concluded that treating of silver oxide has significant and positive effect on the synthesis of stevioside and RiboidiosideA.

**Chart 1**. Results of expression analysis of the *CPPS*, *GGDPS* and *HDS* genes by Real-time PCR. Error bars are based on the standard error.

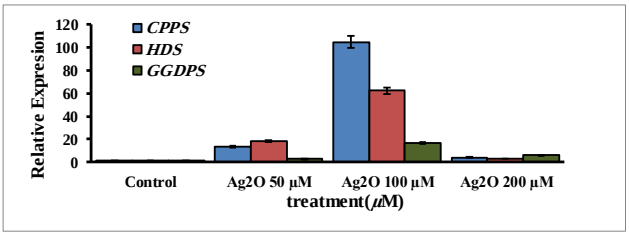
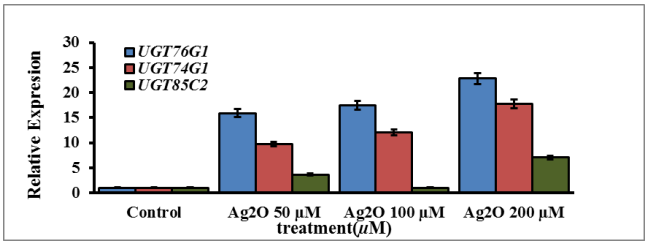



**Chart 2**. Results of expression analysis of the *UGT85C2*, *UGT76G1* and *UGT74G1* genes by Real-time PCR. Error bars are based on the standard error.




### HPLC Analysis

The peak shown in HPLC for stevioside, RiboidiosideA, indicates the increase in the production of these active substances under the influence of silver oxide treatment (Figure 3).
As shown in Figures 2 and 3, the first peak is stevioside and the second peak is RiboidiosideA. Generally, according to the comparison of the standard peaks and silver oxide treatment, it has been shown that the production of stevioside was higher than that of RiboidiosideA, which is evident with regard to the production pathway of RiboidiosideA.

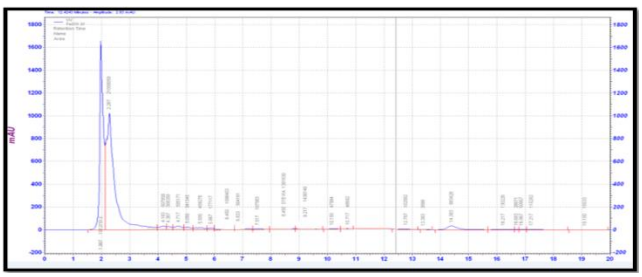


**Figure 2**. HPLC analysis chromatogrm for the standard solution. The first peak represents Stevioside and the second peak represents RebudiosideA.

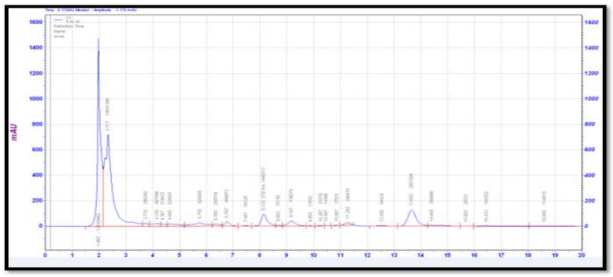


**Figure 3**. HPLC analysis chromatogram of silveroxide: the first peak represents Stevioside and the second peak represents RebudiosideA.

## Biochemical Analysis

As shown in the table 3, treatment of the samples with silver oxide, at concentrations of 100 and 200 µm, have a more positive effect on the levels of chlorophyll, protein, carbohydrates and carotenoids. Also, the results of Duncan test at 0.05% showed a significant difference between the concentration of 50, 100 and 200 µm.

**Table 3**. Silver oxide impact on chlorophyll, carotenoid, carbohydrate and protein contents.

| treatment | chlorophyll (mg/ml) | carotenoid (mg/ml) | protein (mg/g.D.W) | carbohydrate (mg/g.D.W) |
|---|---|---|---|---|
| Control | 15.6±3 | 4.43±2 | 73±6 | 230±12 |
| Ag2O 50 µM | 19.1±8 | 4.36±3 | 102±4 | 274±11 |
| Ag2O 100 µM | 19.19±5 | 4.76±3 | 105±5 | 289±14 |
| Ag2O 200 µM | 20.09±8 | 5.62±2 | 108±3 | 297±13 |
| Positive and negative numbers are related to standard deviation. | | | | |

It is known that, the transcription level of the genes involved in the pathway of ethylene signals is correlated with the defense response of the plants (Kaveh et al., 2013). Metal oxides as an inhibitor of the activity of ethylene, through the coating of the receptors, inhibits the ethylene signal pathway and affects the process of the metabolism and the expression of the associated genes (Krizkova et al., 2008). With regards to the increased level of the expression in the examined genes, it can be concluded that likely silver oxide is involved as a coating on ethylene receptors, which inactivates them. This reaction could lead to changes in the metabolic processes and expression RebaudiosideA. Similar changes have been reported in the metabolic processes of proteins in sunflowers (Krizkova et al., 2008).

The results of HPLC showed that the treatments of silver oxide leads to an increase in the production of Stevioside and RiboidiosideA in stevia. In addition, it was found that the pathway reaches closer to production of RiboidiosideA in the pathway, the activity of enzymes, is increased. Also, the amount of gene expressions synthesizing these enzymes has increased and eventually more RiboidiosideA would be expected. This uptrend suggests that treating the samples with silver oxide would have a significant effect on the expression of these genes. Also, this research suggests that the treatment with silver oxide up-regulates the amount of pigments, chlorophyll content and plant growth, which in turn could reduce the negative effects of the environmental stresses.

## References


1. Arnon A. N. (1967) Method of extraction of chlorophyll in the plants. Agronomy Journal 23：112-121.
2. Bradford M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantites of protein utilizing the principles of protein dyebinding. Analytical Biochemistry Journal 72: 248-254.
3. Bridel M. and Lavielle R. (1931) Sur le principe sucre des feuilles de kaa-he-e (*Stevia rebaundiana B*). Academie des Sciences Paris Comptes Rendus 192: 1123-1125.
4. Chaturvedula V. S. P. Upreti M. and Prakash I. (2011) Structures of the novel α-glucosyl linked diterpene glycosides from *Stevia rebaudiana*. Carbohydrate Research Journal 346(13): 2034-2038.
5. Dacomi E., Mobrien C., Scannell A. G. M. and Arendt E. K. (2005) Evaluation of antioxidant activity of some plant extracts and their application in biscuit. Food Chemistry Journal 90: 118-121.
6. Dibrov P., Dzioba J., Gosink K. K. and Hase C.C. (2002) Chemiosmotic mechanism of antimicrobial activity of Ag+ in Vibrio cholerae. Antimicrobial Agents and Chemotherapy 46(8): 2668-2670.
7. Jain P. K., Huang X. Elsayed A. H. and Elsayed M. A. (2008) Noble metals on the nanoscale optical and photothermal properties and some applications in imaging, sensing, biology and medicine. Accounts of Chemical Research 41: 1578-86.
8. Jeeva K., Thiyagarajan M., Elangovan M., Geetha N. and Venkatachalam P. (2014) Caesalpinia coriaria leaf extracts mediated biosynthesis of metallic silver nanoparticles and their antibacterial activity against clinically isolated pathogens. Industrial Crops and Products Journal 52: 714-20.
9. Jia Z., Sun H. and Gu Q. (2013) Preparation of Ag nanoparticles with triethanolamine as reducing agent and their antibacterial property *Colloids* and *Surfaces A.* Physicochemical and Engineering Aspects 419: 174-9.
10. Jiang H., Moon K. S., Li Y. and Wong C. P. (2006) Surface functionalized silver

nanoparticles for ultrahigh conductive polymer composites. Chemistry of Materials 18: 2969-73.
11. Kailasam S. (2011) Quantification of stevioside and RebaudiosideA in *Stevia rebaudiana* Bertoni leaves using the Agilent 1260 Infinity LC. Agilent Technologies, 1-4.
12. Kaveh R., Li Y. S., Ranjbar S., Tehrani R., Brueck C. L. and Van Aken B. (2013) Changes in *Arabidopsis thaliana* gene expression in response to silver nanoparticles and silver ions. Environmental science and technology 47(18): 10637-10644.
13. Krizkova S., Ryant P., Krystofova O., Adam V., Galiova M., Beklova M. and Liska M. (2008) Multi-instrumental analysis of tissues of sunflower plants treated with silver (I) ions–plants as bioindicators of environmental pollution. Sensors 8(1): 445-463.
14. Kichtenthaler H. k. and Wellburn A. R. (1983) Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvent. Biochemical Society Transactions 603:591-593.
15. Kumar H., Kiran K., Suphla B. G., Vijay K. K. and Sanjay K. (2012) A comprehensive analysis of fifteen genes of steviol glycosides biosynthesis pathway in *Stevia rebaudiana* (*Bertoni*). Gene Journal 492 (1):276-284.
16. Larionov A., Andreas K. and William M. (2005) A standard curve based method for relative real time PCR data processing. BMC bioinformatics 6 (1):1.
17. Livak K. J. and Thomas D. S. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods Journal 25 (4):402-408.
18. Lubick N. (2008) Nanosilver toxicity: ions, nanoparticles or both? Environmental Science Technology 42(23): 8617-8617.
19. Mafune F., Kohno J. Y., Takeda Y., Kondow T. and Sawabe H. (2000) Structure and stability of silver nanoparticles in aqueous solution produced by laser ablation. Journal of Physical Chemistry B. 104: 8333-7.
20. Pereira J. S. and Chaves M. M. (1993) Plant water deficits in mediteranian ecosystems. Water Deficits and Plant Growth. Eds. Plant response from cell to community. vol. IV.pp. 237-251. Academic Press NewYork.
21. Rita E. M. H. (1997). Stevia natures sweetener. Woodland publishing, Inc 160: 1-29.
22. Soejarto D. D. (2002) Botany of Stevia and *Stevia rebaudiana*. Stevia: the genus Stevia. Medicinal and aromatic plants: industrial profiles 19: 18-39.
23. Tadhani M. B., Patel V. H. and Subhash R. (2007) In vitro antioxidant activities of *Stevia rebaudiana* leaves and callus. Journal of Food Composition and Analysis 20(3): 323-329.

Research Article

# MicroRNA Binding Site Polymorphisms are Associated with the Development of Gastric Cancer

Jamal Asgarpour[1], Jamshid Mehrzad[* 2], Seyyed Abbas Tabatabaei Yazdi[3]

[1]Department of Genetic, Islamic Azad University, Neyshabur, Iran
[2] Department of Biochemistry, Islamic Azad University, Neyshabur, Iran
[3] Department of Pathology, Mashhad University of Medical Sciences, Mashhad, Iran

Received 20 November 2017 Accepted 15 December 2017

## Abstract

MicroRNAs (miRNAs) binding to the 3'-untranslated regions (3'-UTRs) of messenger RNAs (mRNAs), affect several cellular mechanisms such as translation, differentiation, tumorigenesis, carcinogenesis and apoptosis. The occurrence of genetic polymorphisms in 3'-UTRs of target genes affect the binding affinity of miRNAs with the target genes resulting in their altered expression. The current case-control study of 100 samples (50% cancer patients and 50% health persons as control) was aimed to evaluate the genotyping of microRNA single nucleotide polymorphisms (SNPs) located at the binding site of the 3'-UTR of C14orf101 (rs4901706) and mir-124 (- rs531564) genes and their correlations with gastric cancer (GC) development. The Statistical analysis results indicated the significant association of AG development risk with the SNPs located in the 3'-UTR of C14orf101 and mir-124. It could be concluded from the results that these genes are associated with the gastric adenocarcinoma.

***Keywords*:** Carcinogenesis, C14orf101 polymorphism, Mir-124, Gastric cancer development

## Introduction

Gastric cancer is regarded as one of the most common malignancies in the world and is the second leading cause of death around the world. According to the World Health Organization (WHO), around one million people in the world were diagnosed with gastric cancer in 2012, and about 70% live in developing countries, especially in eastern Asia, accounting for 6.8% of all cancers. About 50% of gastric cancers are malignant. There are also 723,000 deaths from gastric cancer every year, equivalent to 8.7% of all deaths (Abediankenari et al., 2013).
Considering the high prevalence of gastric cancer in Khorasan Razavi province of Iran and its genetic variations with other regional countries, the current study has been designed to conclude the effectiveness of GC treatment and prevention in the region (Barni et al., 1988). In this study, genomic DNA of 50 gastric cancer patients and 50 healthy controls was extracted from paraffin blocks in gastric. Thereafter, genotyping of C14orf101 and mir-124 genes was performed using PCR-RFLP and statistical analysis using Medcalc software.

The alternative name of C14orF101 is TMR260 (NCBI gene number; rs4901706) which is responsible to translate a transmembrane protein, TMEM260, located on the short arm of chromosome 14 (Compare et al., 2010).The Mir-124 gene (NCBI gene number; rs531564) is located on the short arm of chromosome 8. In many studies, the altered expression of Mir-124 has been observed having no effect on the differentiation of the neuronal cells but affects the types of cancers, especially gastrointestinal (GI) cancers (Tepes, 2009).
The current study was aimed to evaluate the association of MicroRNA binding sites polymorphisms with the development of gastric cancer. For this purpose, 50 cancer patients were included in this study along with the 50 normal healthy persons as control.

### Materials and Methods

Case-control study was designed to investigate the relationship between miRNA polymorphisms in the 3'-untranslated regions (3'-UTRs) of C14orf101 and mir-124 genes of

Corresponding authors E-mail:
* *mehrzadjam@yahoo.com*

the under treatment gastric cancer patients at Ghayem Hospital, Mashhad, Khorasan Razavi, Iran. 50 confirmed patient samples of 2017 (5 women and 45 males aged 50-65 having gastric adenocarcinoma) were taken from pathology department after consultation with the statistics expert in the form of paraffin blocks. Along with the cancer patients samples, 50 paraffin blocks of healthy subjects (those whose pathology results are not difficult and non-cancerous) were also taken as control. Paraffin tissues were dehydrated and genomic DNA was extracted as per kit manufacturer's instruction (ParsTus DNA extraction kit, Cat No.A101211).

## Primers Used

The primers were synthesized by Bioneer Company, South Korea as described in Table 1 (Li et al., 2012).

**Table 1**. Primers Used

| Gene | RS | Primer Sequences | Enzyme | PCR Product Size |
|---|---|---|---|---|
| c140f101 | Rs - 4901706 | Forward:5-CTGAAGTG CTGTTATCG GAAACC-3<br><br>Reverse :5-GCCCTTAC ACAGTTCA TGACCAA-3 | Bsp143 (Sau3) | 494= 314+ 180 |
| mir-124 | Rs-531364 | Forward: 5-CATTGTCTG TGTGATTG GGGGA-3<br><br>Reverse : 5-AAACACAG TCACGGAG GAAGGTG-3 | Alw26I (BsmA) | 682= 500+182 |

To determine the precise binding locus of C14 orf101 and mir-124 genes, their sequences were first determined by NCBI and primers were designed using GeneRunner software and their specificity were determined via BLAST protocol. NEB Cutter software was used to determine the limited enzymes for RFLP reactions.

PCR reaction were performed for locus of C14 orf101 and mir-124 genes according to the PCR amplification protocol shown in table 2. All PCR amplification reactions were conducted for 36 cycles at 3 phases (denaturation, annealing and extension)

**Table 2**. PCR amplification protocol

| PCR Amplification Protocol for c14orf101 | | | |
|---|---|---|---|
| **Denaturation** | 94°C | 5 min | P1 |
| | 94°C | 30 s | Cycles: P2+36 |
| **Annealing** | 64°C | 40 s | |
| **Extension** | 72°C | 40 s | |
| | 72°C | 10 min | P3 |

| PCR Amplification Protocol for Mir -124 | | | |
|---|---|---|---|
| **Denaturation** | 94°C | 5 min | P1 |
| | 94°C | 30 s | Cycles: P2+ 36 |
| **Annealing** | 65°C | 40 s | |
| **Extension** | 72°C | 40 s | |
| | 72°C | 10 min | P3 |

## Result

### Evaluation of PCR by Electrophoresis

RFLP-PCR is a method that has been tested for PCR production in the presence of limited enzymes and the size of the components created using gel electrophoresis was checked alongside the marker (Lerner et al., 1958).

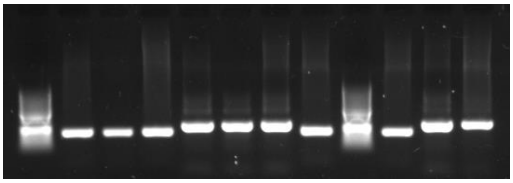


**Figure 1**. Agarose gel 1% extruded genomic DNA of the patient

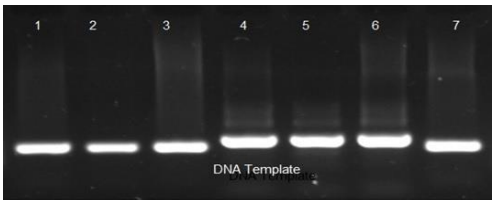



**Figure 2**. Agarose gel 1% extruded genomic DNA from healthy people

### Results of Polymerase Chain Reaction (PCR) Promoter of C14 orf101 and Mir-124 Genes:

After examining the quality of DNA extracted from the control and patients, polymerase chain reaction (PCR) was performed based on the protocol. PCR product of C14orf101 and mir-124 has been yield as a 500-bp and 700-bp fragment, as shown in Figure 3 and 4. These fragments contain the

region in which the desired polymorphism occurs.

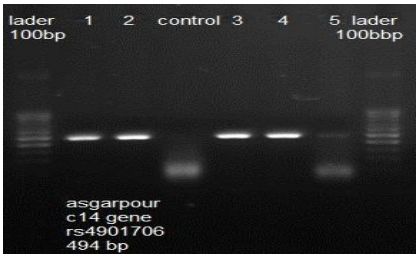



**Figure 3**. Agarose gel electrophoresis of the PCR product of C14orf101 gene.

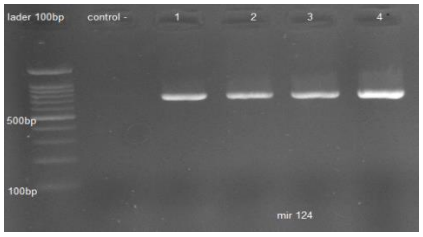
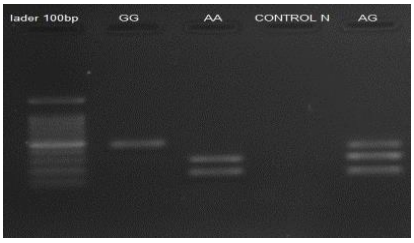



**Figure 4**. Agarose gel electrophoresis of the PCR product of mir-124 gene

**Polymerase Chain Reaction-restriction Fragment Length Polymorphism (PCR-RFLP)**

The primers sequences and the restriction enzyme for analysis of of C14orf101 and mir-124 gene polymorphisms were described in table 1. Figure 5 shows results from PCR-RFLP.

A




B

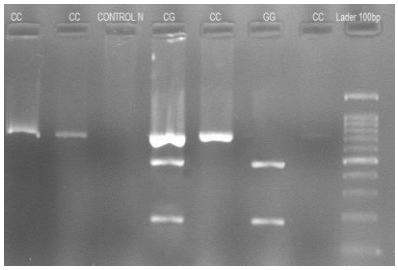



**Figure 5**. RFLP of C14orf101 (Rs -4901706), and mir-124 (Rs-531364). A: For C14orf101 Rs -4901706 gene polymorphism, the wild type homozygous alleles yielded a 494-bp products, while the mutated type homozygous and heterozygous alleles yielded a 184-314-bp and 182-314-494 bp respectively (B) For mir-124 (Rs-531364) gene polymorphism, the wild type homozygous alleles yielded a 682-bp product, while the mutated type heterozygous and homozygous alleles yielded 182-500-682 bp and 182-500bp products respectively.

**Statistical analysis**

Table 3 showed the association between the SNPs and cancer risks. For mir-124, patient carrying homozygous CC genotype are 33%. We have been found that the heterozygote CG and homozygote GG are 60% and 7%. Results for control group of the mir-124 gene is described in Table 3.

**Table 3.** Distribution of different genotype for mir-124 gene in cases and controls.

| | **Number of patients (45)** | | |
|---|---|---|---|
| | Genotype | Frequency | Number of Genotypes |
| Wild type homozygous genotype | CC | 15 | 33% |
| Wild type heterozygote genotype | CG | 27 | 60% |
| Mutant type Homozygote genotype | GG | 3 | 7% |
| | **Number of Controls (48)** | | |
| | Genotype | Frequency | Number of Genotypes |
| Wild type homozygous genotype | CC | 26 | 54% |
| Wild type heterozygote genotype | CG | 22 | 46% |
| Mutant type Homozygote genotype | GG | 0 | 0 |

Relationship between genotype of mir-124 polymorphism and gastric cancer diseases has been investigated with Medcalc software. Difference in genotypic frequency between control and patient groups has been studied by Chi-square test.

The value of Chi- square=6/371 with p-value=0.0414 has been obtained. Therefore, according to p-value≤0/05, there is a significant difference in the distribution of the desired polymorphism between the control and patient groups.

In analyses of c140f101 gene polymorphism, we have been found patient carrying homozygous GG, heterozygote GA and homozygote AA genotype are 60%, 30% and 10%. Also, results for control group of the mir-124 gene is described in Table 4.

**Table 4.** Distribution of different genotype for c140f101 gene in cases and controls.

| | | | |
|---|---|---|---|
| **Number of patients (45)** | | | |
| | Genotype | Frequency | Number of Genotypes |
| Wild type homozygous genotype | GG | 27 | 60% |
| Wild type heterozygote genotype | GA | 14 | 30% |
| Mutant type Homozygote genotype | AA | 4 | 10% |
| **Number of Controls (48)** | | | |
| | Genotype | Frequency | Number of Genotypes |
| Wild type homozygous genotype | GG | 43 | 90% |
| Wild type heterozygote genotype | GA | 5 | 10% |
| Mutant type Homozygote genotype | AA | 0 | 0 |

The relationship between genotypic polymorphism in Table 9 and gastric cancer was assessed by using the Medcalc software. Chi-square test was used to determine the genotypic frequency between two control and patient groups. We obtained Chi-Square= 811.11 with p-value=0.0027. Therefore, considering p-value <0.05, there is a significant difference in the distribution of desired polymorphism between the control and patient groups.

## Discussion

Gastric cancer is one of the most common cancers and secondary causes of cancer deaths in the world. Usually, the risk of developing gastric cancer is higher among people with poor nutrition (Babaei et al., 2010). Immigration is one of the most significant factor associated with the risk of stomach cancer indicating that lifestyle factors such as nutritional habits etc are the determinants of stomach cancer (Jemal et al., 2010). It also has been proven that genetic background and polymorphism of alleles in different races are associated with cancer progression. Polymorphism of C14orf101 and mir-124 genes has also been observed to be linked with gastric cancer malignancy (Lin et al., 2015). Considering this relationship (Siewert and Maruyama, 1995), the current research was designed in Mashhad city of Khorasan Razavi with the clinical assistance of Ghayem Hospital. In a similar study performed in 2014 by Mangan Wang and Ping in China, mir-124 genus with rs531564_was found in gastric cancer patients and absent in healthy persons. Therefore, according to the research carried out, the results of the current study are in consistence with the previous published work but discussing the gene behavior in Khorasan Razavi population.

## Conclusion

The results of this study showed that the polymorphism of the C14orf101 and mir-124 genes are related to gastric adenocarcinoma and screening of the polymorphisms of this gene could be an important factor in the prognosis of disease, prevention of disease progression, as well as the use of appropriate therapeutic approaches to increase longevity and improve quality life in patients with gastric cancer (Lin et al., 2015).

## Reference:

1. Abediankenari S., Shokrzadeh M., Aminjan H. H., Nasri N. and Alizadeh A. (2013) Evaluation of caspase3 and 9 gene polymorphisms in gastric cancer patients in Mazandaran province: a brief report. Tehran University Medical Journal 71.
2. Babaei M., Pourfarzi F., Yazdanbod A., Chiniforush M. M., Derakhshan M. H., Mousavi S. M., Samadi F. and Rahimi G. (2010) Gastric cancer in Ardabil, Iran--a review and update on cancer registry data. Asian Pacific Journal of Cancer Prevention (APJCP) 11:595-599.
3. Barni S., Lissoni P., Crispino S., Cattaneo G., Rovelli F., Fumagalli G. and Tancini G. (1988) Neuroimmunomodulation in cancer patients: correlations between melatonin and beta-endorphin blood levels and T helper/suppressor ratio. The International Journal of Biological Markers 3:82-86.
4. Compare D., Rocco A. and Nardone G. (2010) Risk factors in gastric cancer. European Review for Medical and Pharmacological Sciences 14:302-308.
5. Jemal A., Siegel R., Xu J. and Ward E.

(2010) Cancer statistics, 2010. CA: A Cancer Journal for Cinicians 60:277-300.
6. Lerner A. B., Case J. D., Takahashi Y., Lee T. H. and Mori W. (1958) Isolation of melatonin, the pineal gland factor that lightens melanocyteS1. Journal of the American Chemical Society 80:2587-2587.
7. Li Y.-L., Tian Z., Zhang J.-B. and Fu B.-Y. (2012) CDH1 promoter polymorphism and stomach cancer susceptibility. Molecular Biology Reports 39:1283-1286.
8. Lin X., Zhao Y., Song W.-m. and Zhang B. (2015) Molecular classification and prediction in gastric cancer. Computational and Structural Biotechnology Journal 13:448-458.
9. Siewert J. and Maruyama K. (1995) "What's new in gastric cancer?"—Introduction. World Journal of Surgery 19:483-483.
10. Tepes B. (2009) Can gastric cancer be prevented. Journal of Physiology and Pharmacology 60:71-77.

*Journal of Cell and Molecular Research (2017) 9 (2), 78-83* DOI: 10.22067/jcmr.v9i2.64586

Research Article

# Identification of Non-Ribosomal Peptide Synthetase Modifications Involved in Surfactin Production and Quorum-sensing Operon of *Bacillus subtilis* MJ01 Isolated from Oil-Contaminated Soil

Touraj Rahimi[1], Ali Niazi[1*], Seyed Mohsen Taghavi[2], Esmaeil Ebrahimie[1,3], Shahab Ayatollahi[4], Tahereh Deihimi[1]

[1] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*
[2] *Department of Crop Protection, Shiraz University, Shiraz, Iran*
[3] *Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, Australia*
[4] *School of Chemical & Petroleum Engineering, Sharif University of Technology, Tehran, Iran*

Received 23 March 2017 Accepted 3 July 2017

## Abstract

The isolation of native microorganism that produced biosurfactants in order to oil pollutants bioremediation and hydrophobic oil hydrocarbons availability inter soil texture has become important issues in bioremediation technology. Surfactin is one of the biosurfactants with more application that produced by *Bacillus subtilis* strains could overcome these problems. Thus in this study, we investigated operon which involved in surfactin biosynthesis and its regulator comQXPA operon due to a high level of surfactin biosynthesis by *B. subtilis* MJ01 isolated from oil contaminated soil based on comparative genomics approaches. Surfactin operon localized and compared among six genomes of close relative strains and MJ01 indicated that missense point mutations on genes of surfactin operon were existence. These mutations affected NPRS protein AMP-binding domain that responsible to bind amino acid to correct the situation on surfactin peptide ring. It seems that lack of hemolytic and anti-microbial function of MJ01 surfactin was due to the creation of missense mutation and modifications in the surfactin biosynthesis NPRS enzyme structure. Moreover, *srf* genes expression regulated by comQXPA quorum sensing operon. MJ01 Quorum sensing operon rearrangement showed that part of the *comQ* gene was extended into *comX* gene and these genes had overlap region. Results suggested that in MJ01 genome has been occurred specific combination of *QS* genes organization. Despite high similarity of three genes *comQXP* among MJ01 with BEST7613 and other subtilis strains group, *comA* gene showed high identity with spizizenii strains group.

***Keywords*:** *Bacillus subtilis*, Srf operon, Quorum sensing, *ComQXPA*, Surfactin

## Introduction

Success in technologies of bio remediation is due to the ability of native microbial populations; so that, application of non-native micro-organisms in the alien environment will lead to failure of bio remediation project. Native adapted microbial populations show higher biodegradation potential in a shorter time in the presence of hydro-carbonated pollution. Therefore, in several studies, separation of native microorganisms with bio remediation power from oil and hydrocarbon polluted environment has been used for elimination of pollution (Patowary et al., 2016). Several reports have been offered about separation of various bacteria variants such as *Pseudomonas* and *Bacillus subtilis* from oil reserves or oil polluted soil; which show that these alive specious can be used for bio remediation or production of bio surfactant for reducing or eliminating reducing or eliminating oil pollutions (Gao et al., 2016; Pereira et al., 2013).

One of the main challenges in bio remediation is hydrophobic nature of oil hydrocarbon compounds; this challenge makes these compounds out of the reach of degrading bacteria. Some of the bacteria have gained the ability of secretion of enzymes and secondary metabolites, which aid in the decomposition process during their mutation. These microbes have the potential of producing bio surfactant that is helpful in Solution strengthening of hydrophobic hydrocarbons and increase the accessibility of microbes to the hydrocarbons (Ghazali et al., 2004). Some of the most crucial and applicable bio surfactants, which have been produced by bacteria, are lipo-peptide compounds. The cyclic peptide compounds attached to an

Corresponding authors E-mail:
*Niazi@shirazu.ac.ir

amphiphilic fatty acid chain have bio environmental and industrial applications (Shaligram et al., 2016). These important biological and industrial compounds include a wide range of materials such as terpenes, bacteriocin, lipo-peptide and many other compounds (Shaligram et al., 2016).

Surfactin is one of the most effective bio surfactants, which has different pharmacological activities such as anti-microbe, anti-virus, anti-cancer, anti-fibrinolytic (hemolytic) and surface tension reduction properties. Surfactin lipo-peptide is a lipo heptapeptide containing beta fatty acid-hydroxy with chain length of 13 to 15 (or 16) carbon atoms. So far, a various variant of surfactin have been reported such as lichenysin from *Bacillus licheniformis* or lacidin from *Bacillus pumilus* Some of the surfactin variants have changes in positions amino acids 2, 4 and 7 (Roongsawang et al., 2010). Final lipo peptide can be formed linear, spinal or as a backbone of branched peptide (Jiang et al., 2016). This anti-biotic should synthesis as an ordered linear chain and finally, it should form 7 amino acids and a chain of beta fatty acid-hydroxy and become cyclic. Therefore, extinguishing one of the modules in NPRS gene cluster will result in the production of a lipo peptide with a missing amino acid (Jiang et al., 2016).

According to application of *B. subtilis* bacteria as a plant growth stimulus, surfactin is necessary for formation of biofilm in producer cells and its accumulation around plant root (Aleti et al., 2015), but it prevents formation of biofilm by other bacteria via attachment to cells' surface and their created interference (Roongsawang et al., 2010). It seems that surfactin acts through permeation to the membrane and degrading it; in fact, this action is the reason of its hemolytic, anti-microbial and anti-virus properties (Aleti et al., 2015).

Surfactin has strong surface properties that lead to a reduction of water surface tension from 72 to 27 mN/m in Critical Micelle Concentration (CMC) 25-220 mg/L due to its variants and determined conditions. Substitution effect of Glu1 in the lichenysin variant of surfactin increases its efficiency up to 2 times more than surfactin (Roongsawang et al., 2010).

Biosynthesis gene cluster of surfactin named *srfA* is more than 25 Kb. The operon has an open reading frame (ORF), *srfA*-A, *srfA*-B, *srfA*-C, and *srfA*-Te. The amino acid sequence of first three ORF is homolog with other NRPS; while the last ORF encodes a thioester from putative II type. *SrfA*-A/B/C include 3 moduli and each ORF can be classified to the functional domain type lately (Roongsawang et al., 2010).

Genetic studies show that production of surfactin is under direct effect of 3 chromosome genes of *comA*, *srf* and *sfp*. *Sfp* gene only can be found in surfactin producer races and its transference to surfactin producer races will enhance production of this effective material (Desai and Banat, 1997). *Srf* Operon is required for producing lipo peptide, surfactin anti-biotic, competitive development and efficient sporulation; it activates by Com QXPA Quorum Systems in *B. subtilis* (Guan et al., 2016). Operon operation mechanism of the quorum is through secreting ComX molecules signaling and their accumulation outside the bacteria cells. These ComX molecules attach to ComP receivers on the bacteria membrane and activate them. Activation of ComP leads to phosphorylation of ComA response regulator, which finally can lead duplication from most of the genes. Microarray studies indicate that srfA A-D is under the direct effect of ComQXPA regulon (Oslizlo et al., 2015).

ComX has the interaction effect with histidine kinase attaching to ComP membrane and stimulates the process of auto phosphorylation; it transfers its phosphor to serine residues to respond ComA regulator. Phosphorylated ComA attaches to ComA box (T/GcGG-N4-CCGCA) on the upper side of the srfA promoter as a tetramer, and begins the duplication from srfA. In addition, the mutation in 3 amino acids except aspartate at the end of N- of ComA reduces production of surfactin. These three amino acids may interfere with phosphorylation mechanism. Glucose can stimulate duplication from Com A and consequently increase *srfA* (Roongsawang et al., 2010).

In previous studies that were carried out by Jahanbani et al., (2015) on MJ01 bacteria, it was determined that the bacteria can produce effective material of surfactin at relatively preferable level (1.1 g/lit). Therefore, as operon *srf* is regulated by bacteria quorum system (comQXPA), recognition and understanding the difference between quorum operon of these bacteria with their similar homologous among their close relatives have gained attention.

Moreover, the properties of produced surfactin by MJ01 bacteria such as lack of hemolytic power, anti-bacterial and anti-fungal properties have strengthened the hypothesis that NRPS coding genes, which biosynthesize surfactin, have mutants. Therefore, this study has been carried out to study *srf* operons and quorum through comparing genomics analysis, in order to answer above questions and recognize surfactin biosynthesis mechanism and also the competition among these bacteria.

## Materials and Methods

Bacteria genome of *B. subtilis* has been interpreted and sequenced in the laboratory of biotechnology Institute of Shiraz University. This genome with accessibility Number CP018173.1 is available in NCBI genome data base. Alignment, visibility, and manipulation of sequences have been evaluated by CLC genomic Workbench Software V.09. Genome interpretation data have been used in an online platform of MicroScope (Vallenet et al., 2012; Vallenet et al., 2009).

To identify producer operons of secondary metabolites such as surfactin an online tool named AntiSMASH with the address of http://antismash.secondarymetabolites.org/ was used. AntiSMASH is a predictor and comparing tool, which is used for coding the zones of secondary metabolites, bacteriocin and non-ribosomal lipo-peptide antibiotics by bacteria (Weber et al., 2015).

## Results

### Surfactin Biosynthesis Operon

Surfactin coding operon in the position of 1484194 to 1458804 locates on MJ01 genome. This operon is similar to gene cluster existing in the data base of 86% in terms of their gene sequences (Fig. 1). Consideration of homologs gene cluster of this operon in relative bacteria indicates that this operon has the most similarity (97%) with its homologs bacteria in variants of *B. subtilis* subsp.spizizenii W23and ATCC 6633. The homologs similarities between these operons are 91 and 78% at variants of 168 and spizizenii TU-B-10 (the closest relative (fig2). Based on the data that achieved from AntiSMASH tool and interpretation of MJ01 genome, the positions of SrfA operon and *sfp* gene were determined on the genome. This operon has 4 genes *srfA-ABCD* at positions of 1458804 to 1484950. *Sfp* gene locates in 5017 open pair which is in the lower side of surfactin biosynthesis operon (Table 1).

**Table 1**. The position of surfactin NRPS biosynthesis operon and sfp gene on MJ01 genome

| Gene name | Start | Final | Branch | Length | Fs |
|---|---|---|---|---|---|
| ***srfA*-A** | 1458804 | 1469567 | + | 10764 | F1 |
| ***srfA*-B** | 1469580 | 1480331 | + | 10752 | F2 |
| ***srfA*-C** | 1480367 | 1484194 | + | 3828 | F3 |
| ***srfA*-D** | 1484222 | 1484950 | + | 729 | F4 |
| ***sfp*** | 1489293 | 1489967 | - | 675 | F5 |

Fs: Functions
F1 = Surfactin synthetase
F2 = Surfactin synthetase
F3 = Surfactin synthetase
F4 = Surfactin synthetase
F5 = 4-phosphopantetheinyl transferase

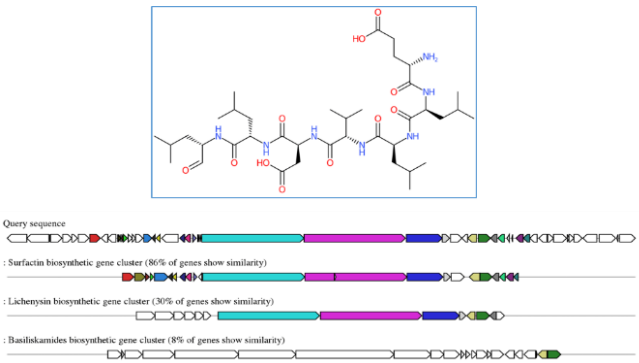



**Figure 1.** Gene cluster biosynthesizing surfactin lipo-peptide and its predicted structure at variant MJ01 and its comparison with gene cluster of other bacterial variants existing in the data base

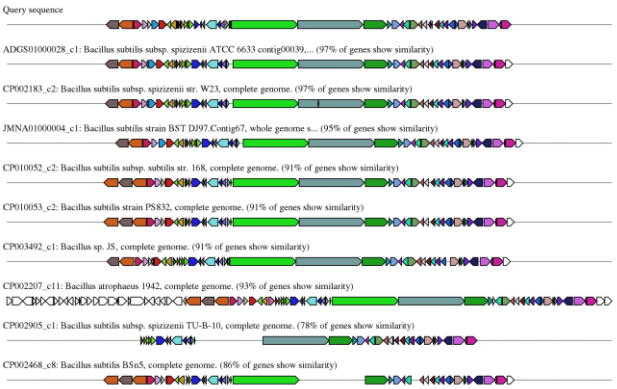



**Figure 2.** Gene cluster comparison between MJ01 and other bacterial variants which biosynthesize surfactin lipo-peptide. Preserved synteny zones are observable in comparison with aimed and considered sequences. Different genes are indicated by different colors and homologs genes are shown by similar colors.

The nucleotide comparison between surfactin operon genes of MJ01 with other relative variants of *B. subtilis* showed that point mutation in genes of this operon caused codon change of missense substitutions. Mutation in NRPS enzyme protein caused changes at the functional domain protein and as a result, the structure of amino acids of surfactin lipo-peptide and their efficiency have been changed (Jiang et al., 2016).

Positioning the missense point mutation on surfactin operon genes and its effect on the functional section of NRPS protein showed that

these mutations affect the AMP-binding domain which is responsible for joining correct amino acid to surfactin peptide chain (Fig. 3).

Jiang et al., (2016) studied mutation on surfactin operon genes and creation of changed surfactin peptide chain; they indicated that mutation in Lucien No. 3 and 6 of produced surfactin by *B. subtilis* bacteria caused it to loss its hemolytic power, but its anti-fungal power has increased. On the other hand, the presence of acid aspartic in surfactin peptide chain will cause losing surfactin anti-microbial power. In the other word, the absence of acid aspartic in surfactin peptide chain increases hemolytic and anti-bacterial powers of produced surfactin (Jiang et al., 2016).

Therefore, it seems that according to surfactine hemolytic, anti-microbial and anti-fungal inactivity, the main reason for changes in surfactin efficiency is missense mutation and changes in structure of NRPS enzyme; therefore, the produced surfactin, which has peptide chain, doesn't have common septet amino acids, and so it doesn't show the common hemolytic and anti-bacterial powers.

**Quorum Operon ComQXPA and its Effect on Surfactin Biosynthesis**

Quorum operon includes 4 genes which locate on 75902 to 80088 bp on complementary strands of MJ01 genome. *Com A* gene has been interpreted as two coding sequences, which each encodes a segment of the regulated protein of quorum (QS); it has been interpreted and recognized as a false gene (Table 2).

BLAST search of the entire sequence of the operon in local data base, which included 40 complete genomes of *B. subtilis* strains, showed that this operon with BEST7613 strain had 94.84% similarity and consequently it was similar to *B. subtilis* subsp.*subtilis* str.KCTC 3135 and delta6 variant (94.84%).

Consideration of similarity of this operon genes by local data base indicated that comA gene had the most similarity with spizizenii TU-B-10 strain (99.84%), while, its similarity with BEST 7613 was 98.28%. Other com QXP genes had the most similarity with BEST 7613 strain which were equal to 95.34, 94.05 and 93.72% orderly. Alignment of this operon with homologs operon in spizizenii TU-B-10 strain showed that nucleotide Polymorphism exists as 3 genes of comQXP in two strains with high frequency.

QS operon guides its regulatory activities via ComA through the cell. The result of studies in *B. subtilis* indicates that more than 10% of genomes are under control of quorum sensing network (QS) which is done by duplication factor of ComA (Wolf et al., 2015). Comell and Grossman (2005) found that most of the genes, which are under the effect of ComA, are coding genes of the membrane or secretory proteins or even proteins that are extracellular and involve in the production of membrane products. These processes include competitive development, enzyme digest, production of anti-biotic and exopolysaccharides, metabolisms and fatty acid transference. Their results showed that regulation path of ComX, ComP and ComA has been acted by Pheromone and kinases that are only under the effect of ComA (Comella and Grossman, 2005)

Dogsa et al., (2014) considered quorum system of *B. subtilis* and found that bacteria of these two groups can be divided into two groups in terms of quorum operon organization. The first group includes quorum operon genes that have not overlap, and the second group includes *comQ* and *comX* genes that have overlap. Because of the stop codon mutation, this overlap has caused the development of final C from *comQ* 13-18 amino acid to *comX* (Dogsa et al., 2014).

The study of quorum operon organization of MJ01 showed that this bacterium had overlap operon and some parts of the *comQ* gene have extended to the *comX* gene; based on Dogsa et al. classification *B. subtilis* MJ01 locates in the second group. It seems that overlap of reading framework leads to controlling character duplication and reduction of the requirement for complex regulatory paths (Johnson and Chisholm, 2004). In addition, this kind of overlapping organization in involved genes is common in regulating processes such as Loci that regulate comQXPA genes expression in bacteria (Comella and Grossman, 2005; Dogsa et al., 2014).

As a result, it seems that in MJ01, a unique type of shuffle combination of QS operon genes has occurred. Qua despite the high similarity of three genes (*comQXP*) from Qs operon in the MJ01 bacterium with BEST7613 and bacteria of subtilis group, *comA* gene of this operon is similar with strains of the spizizenii group as well.

Despite MJ01 genome has all the biosynthesis genes for antibiotic and peptide bacteriocins and is able to confirm the presence of resistance genes to the antibiotic, it shows a different appearance in the laboratory media. Therefore, it seems that according to logical relationship of antibiotic biosynthesis operons such as *srfA* operon (surfactin producer) with signaling and logical systems of *comQXPA*, this regulatory system and mutations in NRPS enzyme, which is responsible for surfactin biosynthesis, can be factors for changing the behaves of MJ01.

**Table 2**. Characteristics and positions of quorum operon *comQXPA* on MJ01 bacteria genome

| Gene name | Start | Final | Branch | Length | Fs |
|---|---|---|---|---|---|
| ***comQ*** | 79102 | 80088 | - | 987 | F1 |
| ***comX*** | 78951 | 79118 | - | 168 | F2 |
| ***comP*** | 76627 | 78936 | - | 2310 | F3 |
| ***comA*-1** | 76388 | 76546 | - | 159 | F4 |
| ***comA*-2** | 75902 | 76387 | - | 486 | F5 |

Fs: Functions
F1 = Isoprenyl transferase
F2 = Competence pheromone precursor
F3 = Two-component sensor histidine kinase
F4 = Two-component response quoeume-sensing regulator-part-1
F5 = Two-component response quoeume-sensing regulator-part-2

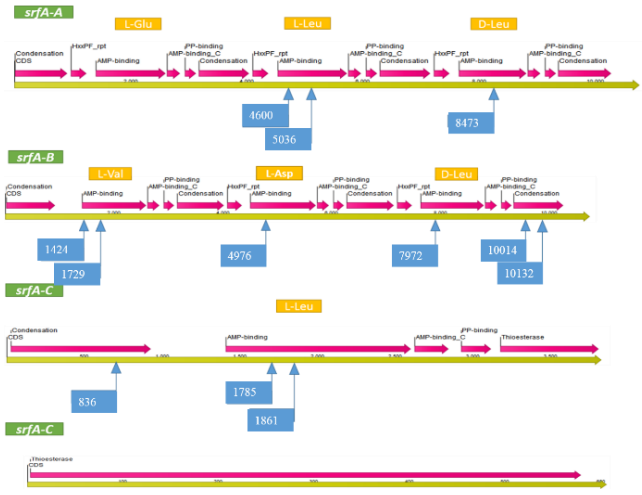



**Figure 3**. The position of the missense mutation on surfactin operon gene of MJ01 genome. The functional domain has been done based on Pfam (pink). A missense mutation (the blue boxes) located on the AMP-binding domain of surfactin biosynthesis enzyme (NRPS). The digits inside blue boxes show nucleotide number from the beginning part of the gene. Involved amino acids of each domain- binding have been mentioned in yellow boxes.

## Discussion and Conclusion

Discovering and annotating coding sequences of PKS and NRPS enzymes in MJ01 genome and its comparison with relative strains and references indicated that although extracted lipo-peptides from this bacterium do not show anti-microbial, anti-fungal and hemolytic activities, there is not any significant difference between that enzymes and bacteria of *B. subtilis* group and their close relatives in terms of operon organizations, the structure of biosynthesis genes of biosurfactants and sequence of nucleoid. Therefore, it seems that enzyme machine that biosynthesis non-ribosomal lipo-peptide in MJ01 bacteria may be the subject to changes or modifications after translation of enzyme protein, as a result it causes their peptide products could not show its effect. Mutation in subunits of enzyme protein domain (NRPS) can substitute amino acids in the structure of lipo-peptide chain (Jiang et al., 2016). This substitution can reduce surfactin activity in *B. subtilis* up to two times. Therefore, the exact study of reasons and effective factors in the appearance of such inefficiency can be a solution for biosynthesis engineering of anti-fungal and anti-microbial lipo-peptides.

Proper growth of this bacterium in culture media showed that it should have high competition and quorum abilities. Although high polymorphism in *comQXPA* genes of *B. subtilis* bacteria which have been separated from soil has been reported previously (Oslizlo et al., 2015), but considering *comQXPA* quorum operon in MJ01 bacteria showed that nucleotide changes caused that *comA* gene has been considered as a false gene in the interpretation of MJ01 genome. Therefore, according to regulating the role of QS operon gene, consideration of gene network relationship under the effect of *comQXPA* and its effect on secretory mechanisms and extracellular lipo-peptides activities of MJ01 bacteria through data of genome expression seems necessary.

## Acknowledgment

This paper is a part of a Ph.D. dissertation that was carried out in biotechnology institute of the agriculture department of Shiraz University; the author is thankful of the University officials for providing the cost of this research.

## References

1. Aleti G., Sessitsch A. and Brader G. (2015) Genome mining: prediction of lipopeptides and polyketides from Bacillus and related Firmicutes. Computational and Structural Biotechnology Journal 13:192-203.
2. Comella N. and Grossman A. D. (2005) Conservation of genes and processes controlled by the quorum response in bacteria:

characterization of genes controlled by the quorum-sensing transcription factor ComA in *Bacillus subtilis*. Molecular Microbiology 57:1159-1174.
3. Desai J. D. and Banat I. M. (1997) Microbial production of surfactants and their commercial potential. Microbiology and Molecular Biology Reviews 61:47-64.
4. Dogsa I., Choudhary K. S., Marsetic Z., Hudaiberdiev S., Vera R., Pongor S. and Mandic-Mulec I. (2014) ComQXPA quorum sensing systems may not be unique to *Bacillus subtilis*: a census in prokaryotic genomes. PloS One 9:e96122.
5. Gao P., Li G., Li Y., Li Y., Tian H., Wang Y., Zhou J. and Ma T. (2016) An exogenous surfactant-producing *Bacillus subtilis* facilitates indigenous microbial enhanced oil recovery. Frontiers in Microbiology 7:186.
6. Ghazali F. M., Rahman R. N. Z. A., Salleh A. B. and Basri M. (2004) Biodegradation of hydrocarbons in soil by microbial consortium. International Biodeterioration & Biodegradation 54:61-67.
7. Guan C., Cui W., Cheng J., Zhou L., Liu Z. and Zhou Z. (2016) Development of an efficient autoinducible expression system by promoter engineering in *Bacillus subtilis*. Microbial Cell Factories 15:66.
8. Jiang J., Gao L., Bie X., Lu Z., Liu H., Zhang C., Lu F. and Zhao H. (2016) Identification of novel surfactin derivatives from NRPS modification of *Bacillus subtilis* and its antifungal activity against *Fusarium moniliforme*. BMC Microbiology 16:31.
9. Jhnson Z. I. and Chisholm S. W. (2004) Properties of overlapping genes are conserved across microbial genomes. Genome Research 14:2268-2272.
10. Oslizlo A., Stefanic P., Vatovec S., Beigot Glaser S., Rupnik M. and Mandic-Mulec I. (2015) Exploring ComQXPA quorum-sensing diversity and biocontrol potential of Bacillus spp. isolates from tomato rhizoplane. Microbial Biotechnology 8:527-540.
11. Patowary K., Patowary R., Kalita M. C. and Deka S. (2016) Development of an efficient bacterial consortium for the potential remediation of hydrocarbons from contaminated sites. *Frontiers in Microbiology* 7:1092.
12. Pereira J. F., Gudiña E. J., Costa R., Vitorino R., Teixeira J. A., Coutinho J. A. and Rodrigues L. R. (2013) Optimization and characterization of biosurfactant production by *Bacillus subtilis* isolates towards microbial enhanced oil recovery applications. Fuel 111:259-268.
13. Roongsawang N., Washio K. and Morikawa M. (2010) Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. International Journal of Molecular Sciences 12:141-172.
14. Shaligram S., Kumbhare S. V., Dhotre D. P., Muddeshwar M. G., Kapley A., Joseph N., Purohit H. P., Shouche Y. S. and Pawar S. P. (2016) Genomic and functional features of the biosurfactant producing Bacillus sp. AM13. Functional & Integrative Genomics 16:557-566.
15. Vallenet D., Belda E., Calteau A., Cruveiller S., Engelen S., Lajus A., Le Fèvre F., Longin C., Mornico D. and Roche D. (2012) MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Research 41:D636-D647.
16. Vallenet D., Engelen S., Mornico D., Cruveiller S., Fleury L., Lajus A., Rouy Z., Roche D., Salvignol G. and Scarpelli C. (2009) MicroScope: a platform for microbial genome annotation and comparative genomics. Database 2009.
17. Weber T., Blin K., Duddela S., Krug D., Kim H. U., Bruccoleri R., Lee S. Y., Fischbach M. A., Müller R. and Wohlleben W. (2015) antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Research 43:W237-W243.
18. Wolf D., Rippa V., Mobarec J. C., Sauer P., Adlung L., Kolb P. and Bischofs I. B. (2015) The quorum-sensing regulator ComA from *Bacillus subtilis* activates transcription using topologically distinct DNA motifs. Nucleic Acids Research 44:2160-2172.

*Journal of Cell and Molecular Research (2017) 9 (2), 84-97* DOI: 10.22067/jcmr.v9i2.66166

Research Article

# Systems Biology Analysis of the Key Genes of Surfactin Production in *Bacillus subtilis* MJ01 (Isolated from Soil Contaminated Oil in South of Iran), Spizizenii, and 168 Isolates

Tahereh Deihimi[1], Esmaeil Ebrahimie[1,5]*, Ali Niazi[1], Mansour Ebrahimi [2] , Shahab Ayatollahi[3] , Ahmad Tahmasebi [4] , Touraj Rahimi[1], Moein Jahanbani Veshareh[6]

[1]*Institute of Biotechnology, Shiraz University, Shiraz, Iran.*
[2] *Department of Biology, University of Qom, Qom, Iran*
[3]*School of Chemical & Petroleum Engineering, Sharif University of Technology, Tehran, Iran*
[4]*Department of Crop Production & Plant Breeding, College of Agriculture, Shiraz University, Shiraz, Iran*
[5] *Adelaide Medical School, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, Australia*
[6] *Danish Hydrocarbon Research and Technology Center, Kgs. Lyngby, Denmark*

Received 17 July 2017 Accepted 23 Agust 2017

## Abstract

Applying microorganism in oil recovery has attracted attentions recently. Surfactin produced by *Bacillus subtilis* is widely used industrially in a range of industrial applications in pharmecutical and environmental sectors. Little information about molecular mechanism of suffactin compound is available. In this study, we performed promoter and network analysis of surfactin production genes in *Bacillus subtilis* subsp. MJ01 (isolated from oil contaminated soil in South of Iran), spizizenii and 168. Our analysis revealed that *comQ* and *comX* are the genes with sequence alterations among these three strains of *Bacillus subtilis* and are involved in surfactin production. Promoter analysis indicated that *lrp*, *argR*, *rpoD*, *purr* and *ihf* are overrepresented and have the highest number of transcription factor binding sites (TFBs) on the key surfactin production genes in all 3 strains. Also the pattern of TFBs among these three strains was completely different. Interestingly, there is distinct difference between 168, spizizenii and MJ01 in their frequency of TFs that activate genes involve in surfactin production. Attribute weighting algorithms and decision tree analysis revealed *ihf*, *rpoD* and *flHCD* as the most important TF among surfactin production. Network analysis identified two significant network modules. The first one consists of key genes involved in surfactin production and the second module includes key TFs, involved in regulation of surfactin production. Our findings enhance understanding the molecular mechanism of surfactin production through systems biology analysis.

***Keywords*:** Surfactin production gene, Transcription factor (TF), Promoter analysis, Network analysis.

## Introduction

Finding out alternative technologies to increase oil recovery from oil fields around the world has drawn attention for many years. Recently, using microorganisms in this field is popularized (Shibulal et al., 2014). Hence, identification and characterization of novel strains are in demand. *Bacillus subtilis* has been used as a model organism because of its ability to produce surface active compounds (biosurfactants) with highly desirable properties for oil recovery (Anuradha S, 2010). Surfactin shows a remarkable membrane-active and surface-interface properties with a number of biological activities in health care and biotechnology-based processes.

Surfactin draws biotechnologists attention as a potent candidate drug for the resolution of a number of global issues in medicine (Banat et al., 2010; Cao et al., 2010) , industry (Abdel-Mawgoud et al., 2008; Nitschke and Costa, 2007), and environmental protection (Mulligan, 2009). Consequently, an urge has been formed towards understanding the molecular mechanism, gene network and transcriptomic comparison of genes involved in surfactin production.

The dynamics of surfactin production by *Bacillus subtilis* is still ambiguous because of shortage in information on different levels of functional genomics such as promoter activation. While gene function is the result of interactive between upstream non coding promoter region and downstream coding sequence, most of the studies focus on genes (Deihimi et al., 2012). Undoubtedly, examination of

Corresponding authors E-mail:
* esmaeil.ebrahimie@adelaide.edu.au

upstream sequences of genes with similar expression pattern is very important (Yada et al., 1997). The role of transcription factors in controlling the expression of many genes involved in surfactin production in *Bacillus subtilis* is still in paucity and has not been studied in details. Gene network analysis, especially regulatory gene network, is a strong well developed tool in understanding the central genes (hubs) and gene interactions (Shoemaker and Panchenko, 2007). Data mining tools such as decision trees can be used to associate the result of different decisions (Ebrahimi et al., 2015). Moreover, these computational analytical tools discover function of genes and proteins structures and decipher the interactions of genes and also genes with transcription factors more clear (Pashaiasl et al., 2016a).

Herein, comparison of surfactin genes between *Bacillus subtilis* subsp. MJ01, spizizenii and 168 has been performed. In addition, for the first time, *in silico* promoter analysis and prediction of involved TFs and also network analysis of genes associated with surfactin production carried out in order to open a new avenue in molecular mechanism of producing surfactin by *Bacillus subtilis.*

## Materials and Methods

### Sequence Comparison of Key Genes Involved in Surfactin Production and Interacted Genes in *Bacillus subtilis* subsp. *168, Spizizenii* and *MJ01*

Protein and gene sequences of surfactin producing genes were retrieved from NCBI in *B. subtilis* subsp. 168 genome. Genome sequences of *Bacillus subtilis* subsp. 168 (AC: NC_000964), spizizenii (AC: NC_016047) and MJ01 (AC: CP018173) were also extracted from NCBI. The nucleotide and protein sequences of all genes associated with surfactin production isolated from *Bacillus subtilis* subsp. 168 were compared with *B. subtilis* subsp. spizizenii and MJ01 genome using blastn and tblastn respectively using CLC bio Genome Workbench software.

### Gene Interactions, Promoter Analysis and Comparative TF Activation Patterns of Genes Involved in Surfactin Production

In order to identify genes associated with surfactin production and their interactions, we used STRING server (http://string-db.org) (Szklarczyk et al., 2014). The gene sequence of each protein was obtained from NCBI database. We also used NCBI to find the genomic sequence of *Bacillus subtilis* subsp. spizizenii (NC_016047) and *Bacillus subtilis* subsp. Subtilis str. 168 (NC_000964). The genomic locations of genes involved in surfactin production were identified by CLCbio genomic workbench. The potential promoter regions of these genes for all 3 strains (MJ01, 168 and spizizenii) were extracted by selecting the region between each gene (or operon) and the next gene (or operon) as previously described (Mahdi et al., 2014). BPROM algorithm (http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb) (Lee and Chen, 2002) was used to confirm the presence of promoter an its -10 and -35 sites for genes associate in surfactin production. TFs were predicted for all promoter sequences in all 3 strains (MJ01, spizizenii and 168) using BPROM.

### Network Analysis

The genes related to surfactin production were selected from the STRING online tool (Szklarczyk et al., 2014) with the cut-off criterion of combined score > 0.4. The relationships of the nodes degree ≤ 5 were abandoned. We added the TF identified by our promoter analysis to the list of genes for network construction. Finally, we visualized the network by Cytoscape 3.4.0 (Shannon et al., 2003). clusterONE algorithm (Nepusz et al., 2012) was employed to identify the significant modules . Moreover, the nodes with high degree and interaction were defined as hub proteins in the network. The node degree ≥3 were selected as the threshold.

### Data Mining

To find the pattern in data of genes and promoters, three sets of data generated:

1. PD200Genes: A promoter dataset of 200 genes which has been selected randomly among all *Bacillus subtilis* MJ01 genes and 5 key surfactine production associating gene containing 53 variables of promoters.
2. PD5NGenes: A promoter dataset of 5 genes which has been selected randomly among 200 random selective gene and 5 key surfactin production associating gene containing 53 variables of promoters. The variables were the number of promoters for each gene (numeric variables).
3. PD5BGenes: Again the same dataset as above (promoters of 5 randomly selected genes) from each group created but each feature set as binomial (the presence or absence of each promoter set as Yes/No values).

All three datasets were imported into RapidMiner Studio software separately (RapidMiner 5.0.001, Rapid-I GmbH, Stochumer Str. 475, 44,227 Dortmund, Germany); the type of gene set as target or labels variable and other 53 of features of promoters appointed as regular variables. The

following data mining algorithms applied on dataset: Attribute Weighting Algorithms and Decision Tree Models

**Attribute Weighting Algorithms**

Attribute weighting algorithms identify the most important attributes or features which differ between two groups of target or labeled attributes (Torkzaban et al., 2015). The models use various statistical approaches to perform the analysis. The following attribute weighting models were applied on datasets: weight by information gain, weight by information gain ratio, weight by rule, weight by deviation, weight by chi squared statistic, weight by Gini index, weight by uncertainty, weight by relief, weight by PCA, and weight by SVM. The algorithms definitions have already been described in our previous paper (Pashaiasl et al., 2016a). Weights were normalized into the interval between 0 and 1 to allow the comparison between different methods.

**Decision Tree Models**

Decision tree algorithms provide visual explanation of the most important features through depicting an inverted tree with the most important feature as root and other variables as leaves. Various decision trees including Random Forest, Decision Stump Decision, ID3, CHAID and Random Tree were applied on dataset. Details of each decision tree model have also been presented before (Pashaiasl et al., 2016b). To calculate the performance of decision tree models in predicting the right class of soil or non-soil group10-times cross validation was applied on dataset. This approach divides data into 10 parts and each time train the model with 9 parts and then test the model by the last part and computes the efficiency of it; repeating it for 10 times and gives the mean performance value.

**Result**

**Analysis of Surfactin Production Genes Between *Bacillus subtilis* Strains (168, Spizizenii, and MJ01 Strain)**

Our previous study identified MJ01 has more desirable properties for oil recovery than other strains (data not published). Here we attempted to investigate genes, promoters and transcription factors (TFs) involved in surfactin production of this strain to explain its importance over the other strain. Hence, we investigated the sequence difference of the key genes associated whith surfactin production between *Bacillus subtilis* MJ01, *Bacillus subtilis* subsp. spizizenii (which has 95% similarity with MJ01 in genome sequence) and *Bacillus subtilis* subsp. Subtilis str. 168 (which is reference genome) to explain this priority (Table 1). The comparative result revealed that there is no significant difference in sequence of the surfactin producing genes (*srfA*, *srfAA*, *srfAB*, *srfAD*, *sfp*) between 168 vs spizizenii and MJ01. Specially, no difference was observed between the surfactin producing genes sequence in spizizenii vs MJ01, and it can be concluded that they are similar for their sequences.

**Table 1.** Percentage identity of nucleotide sequence in genes associated with surfactin production among *Bacillus subtilis* 168 vs spizizenii vs MJ01

| Gene name | % identity of 168 vs spizizenii | % identity of 168 vs MJ01 | % identity of spizizenii vs MJ01 |
|---|---|---|---|
| **srfA** | 92 | 92 | 99 |
| **srfAA** | 92 | 92 | 99.5 |
| **srfAB** | 92 | 92 | 99 |
| **srfAD** | 93 | 93 | 98 |
| **Sfp** | 93 | 93 | 99 |

Hence, by gene network analysis, we tried to find the genes associated with the surfactin producing genes. Sequence similarity assay for all of the genes has been identified by all networks, between 168 vs spizizenii and MJ01 revealed that there is no significant variance between spizizenii and MJ01 nucleotide sequence for the genes associate with surfactin producing networks, but *ppsC*, *ppsA*, *comQ*, *comX* and *YndJ* has less than 90% identity among 168 vs spizizenii and MJ01 (supplementary 1). As a result of protein sequence comparison between 168 vs spizizenii and MJ01, *pksM*, *pksJ*, *pksD*, *pksL*, *pksR*, *ppsC*, *ppsA*, *comQ*, *comX* and *yndJ* showed less than %90 identity. In contrast, *comQ* and *comX* revealed significant variance between spizizenii and MJ01 in their protein sequences (Fig 1.) (Supplementary 2.).Oil activates different pattern of TFs among different strains of *Bacillus subtilis* (168, spizizenii and MJ01)

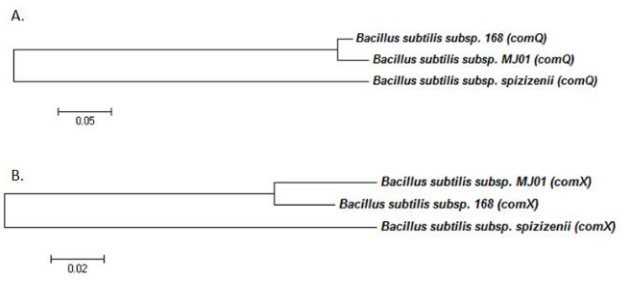



**Figure 1**. Phylogenetic relationship of genes *between Bacillus subtilis* subsp. spizizenii, 168 and MJ01. A. phylogenetic relationships of comQ among *Bacillus subtilis* subsp. spizizenii, 168 and MJ01. B. phylogenetic relationships of comK among *Bacillus subtilis* subsp. spizizenii, 168 and MJ01. The tree was built by using neighbour-joining algorithm.

We sought to identify if there is any difference in TFs capable of binding to genes associate with surfactin production between 3strains of *Bacillus subtilis* (168, spizizenii and MJ01). Generally, the differential TF activation profiles were observed as a result of comparison between 168, spizizenii and MJ01. Among all TFs identified for the promoters of genes associated with surfactin production, the number of *marR* and *ihf* was the same in all three genomes (*Bacillus subtilis* 168, *Bacillus subtilis* spizizeni and *Bacillus subtilis* MJ01).Thirty seven percent difference was observed in the number of TFs of surfactin producing genes promoter by comparing *Bacillus subtilis* MJ01 and *Bacillus subtilis* spizizenii. TFs such as *Lrp*, *argR*, *rpoD*, *purR* and *ihf* have the highest number of TFBSs (at least 5) between all 3 strains. The three highest numbers of TFs in *Bacillus subtilis* 168 gene promoter in surfactin associated genes were *rpoD*, *lrp*, *argR* by 13, 11 and 10, respectively. In *Bacillus subtilis* spizizenii, *rpoD*, *argR* and *purR* with 19, 11 and 8 predicted binding sites were the three highest number of TFs found on promoters of genes involve in surfactin production. The top three number of TFs in *Bacillus subtilis* MJ01 were similar to *Bacillus subtilis* spizizenii but with different numbers (*rpoD*, *argR*, and *purR* with 15, 8 and 8 predicted sites respectively). Our promoter analysis for *srfA*, *srfAA*, *srfAB* and *srfAD* which are the most important genes in surfactin production (Porob et al., 2013) indicates that *rpoH2* has binding site on promoters of these genes in *Bacillus subtilis* subsp. spizizenii but no binding site for *Bacillus subtilis* subsp. MJ01.
Surprisingly, TF analysis revealed that there is a distinct difference between the number of TFs that activate genes associate with surfactin production, between 168 and spizizenii and MJ01 (Table 2).

**Table 2.** Transcription factors (TFs) number of binding sites on the promoter region of surfatin producing genes which were different between all 3 strains (*Bacillus subtilis* 168, spizizenii and MJ01 ) for their TF patterns.

| Gene name | TFs number | | |
|---|---|---|---|
| | **168** | **spizizenii** | **MJ01** |
| **menB** | 2 | 2 | 1 |
| **dhbB** | 7 | 7 | 4 |
| **codY** | 3 | 2 | 0 |
| **ppsC** | 13 | 13 | 4 |
| **ppsA** | 13 | 13 | 2 |
| **dhbF** | 7 | 7 | 4 |
| **ybdZ** | 7 | 7 | 4 |
| **srfAB** | 13 | 10 | 10 |
| **srfAA** | 13 | 10 | 10 |
| **srfAD** | 13 | 10 | 10 |
| **srfA** | 13 | 10 | 10 |

Although the sequences of these genes and their proteins were the same in spizizenii and MJ01, but the number of TFs binding to their promoters are completely different. However, *sfp*, *pksR*, *pksN*, *pksM*, *pksJ*, *pksD*, *comK*, *mecA*, *rapC*, *dltA*, *YndJ* and *pksS* recognized the same TF pattern for spizizenii and MJ01 but completely different for 168. Even *comx* and *comQ* which have different sequence between spizizenii and MJ01, but their TF pattern is completely the same

## Construction of Surfactin Production Gene Regulatory Network in *Bacillus subtilis* subsp. 168

The surfactin production gene regulatory network was generated by combining predicted TFs, target genes and genes associated in surfactin production. *lexA* and *rpoD* with the most edge number among TFs were regulatory hub in this network which play key role in surfactin production. The network of each key surfactin production gene has been constructed (Figure 2.). *Ihf*, *PksM*, *PksD*, *PksL*, *Ybdz* and *arcA* are the joint node between *srfA*, *srfAA*, *srfAB*, *srfAD* and *sfp* network. *Srf* genes and *sfp* have interaction with each other in this surfactin production network which was prospective. *Ihf* and *arcA* are the most common TFs between *srfA*, *srfAA*, *srfAB*, *srfAD* and *sfp* network. *argR2*, *rpoH*, *cytR* and *rpoD* were involved in 80% of key gene networks. The genes involved in surfactin production have enriched functional groups by "Biological Process" term (Table 3).

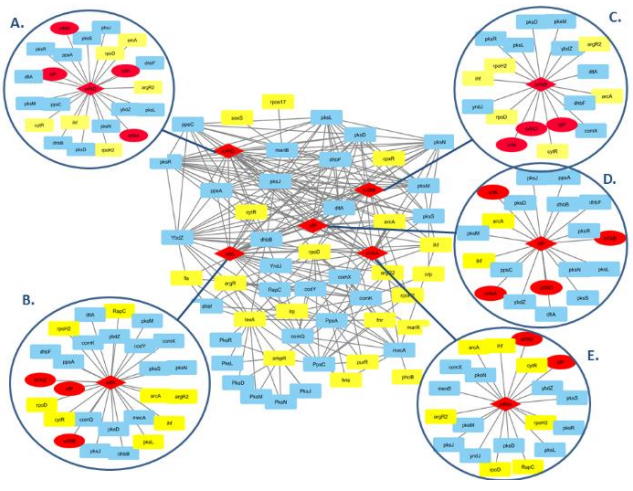



**Figure 2**. Regulatory network of surfactin production. The network has 55 nodes and 262 links. Diamond red nodes represent key surfactin production genes, yellow nodes represents transcription factors (TFs). A., B., C., D. and E. are the subnetwork centred by fAD, srfA, srfAA, sfp and srfAB, respectively.

**Table 3**. Functional classification of gene network of genes involve in surfactin production

| BioProcess | Number of genes involve |
|---|---|
| Antibiotic Biosynthetic Process | 10 |
| Biosynthetic Process | 24 |
| Cellular Biosynthetic Process | 22 |
| Cellular Metabolic Process | 24 |
| Metabolic Process | 28 |
| Transcription DNA template | 10 |
| Regulation of transcription, DNA template | 10 |
| Gene expression | 10 |
| Cellular macromolecule metabolic process | 11 |
| Transcription initiation from bacterial-type RNA polymerase promoter | 2 |
| Transcription from bacterial-type RNA polymerase promoter | 2 |
| primary metabolic process | 11 |
| Positive regulation of transcription, DNA-templated | 3 |

## Subnetwork Analysis of Genes and TFs Involved in Surfactin Production in Reference Genome

Two modules have been constructed using ClusterONE in Cytoscape software. Two modules under the condition of more than 5 nodes were screened and modular significance *P* value less than 0.05 were obtained. Analysis of these modules indicated that module 1 was enriched by the key genes involved in surfactin production. Thirty seven percent of genes consisting module 2 are TFs which directly or indirectly regulate surfactin production (Fig 3).

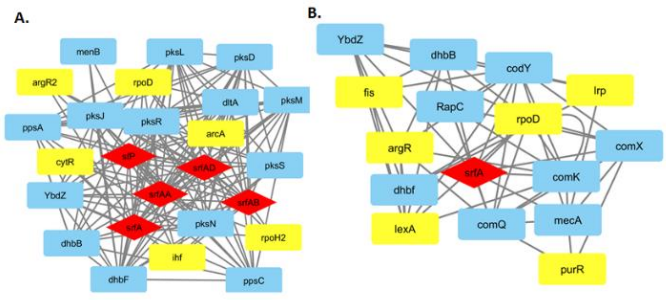



**Figure 3**. The modules identified from the regulatory network of surfectin produvction using ClusterONE. Yellow round rectangular represent transcription factors, and red diamond represent for key surfactin producing genes. A. and B. indicates first and second, respectively.

## Decision Tree and Data Mining Analysis Selected *flHCD, ihf* and Two Members of *rpoD* Groups as the Most Important TF

As mentioned in Materials and Methods, three different datasets of gene promoters were prepared based on the number of samples (in PD200Genes the number of samples were 200; while in two other datasets –PD5NGenes and PD5BGenes – the number of samples was only 5 in each group to balance the number of samples in each group).

### Attribute Weighting Algorithms

PD200Genes: When attribute-weighting algorithms applied on this dataset, *flHCD* attribute was selected as the most important feature by at least 75% of models. Seven algorithms (SVM, Uncertainty, Gini Index, Chi Squared, Rule, Info Gain Ration and Info Gain) generated the highest possible weight of 1.0 to this gene. SVM and Rule algorithms were appointed 1.0 weight to *ihf* and *rpoD17* gene variables; putting them in the next positions. Details of other weights have been shown in Table 4.

### PD5NGenes

When the number of samples in each group balanced and 5 random samples were taken from the main dataset, the attribute weighting algorithms ran on this dataset showed that *ihf* attribute received the best weight of 1.0 by nearly 80% of algorithms; 100% of them gave weights higher than 0.75; confirming the importance of this feature. The other features such as *flHCD*, *rpoD18* and *rpoD17* were also gained more weights here (Table 5).

### PD5BGenes

For this dataset, the presence or the absence of promoters in each gene marked as Yes or No (binomial features created). The results of attribute weighting showed that *ihf* feature gained the best score and marked at the most important feature by 85% of attribute weighting algorithms. *flHCD* and *rpoD18* were among the other features selected by attribute weighting algorithms as second most important (Table 6).

### Decision Trees

Various decision trees were applied on three datasets; the accuracies of each decision tree model calculated based on 10-fold cross validation (dataset divided into 10 equal sets, each time 9 sets used to train the model and tested with the last set, then the model repeated with another 9 and 1 sets and the average of 10 run performances calculated and reported).

**Table 4**. Ten different attribute weighting models applied on all samples showing the most important transcription factors based on various levels of weights assigned by each model.

| SVM | Relief | Uncertainty | Gini Index | Chi Squared | Deviation | Rule | Info Gain Ratio | Info Gain | Attribute | Count 50% | Count 75% | Count 95% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1.0** | .1 | 1.0 | 1.0 | 1.0 | .3 | 1.0 | 1.0 | 1.0 | flHCD | 7 | 7 | 7 |
| **.6** | .1 | .3 | .3 | .3 | .6 | 1.0 | .3 | .6 | rpoD15 | 4 | 1 | 1 |
| **.6** | .0 | .3 | .3 | .3 | .6 | 1.0 | .3 | .6 | rpoD18 | 4 | 1 | 1 |
| **.4** | .7 | .3 | .3 | .3 | .8 | 1.0 | .2 | .7 | ihf | 4 | 2 | 1 |
| **1.0** | .3 | .3 | .2 | .2 | .6 | 1.0 | .2 | .5 | crp | 3 | 2 | 2 |
| **.3** | .2 | .4 | .3 | .4 | .6 | 1.0 | .3 | .6 | purR | 3 | 1 | 1 |
| **.4** | 1.0 | .1 | .0 | .0 | .9 | 1.0 | .0 | .2 | rpoD17 | 3 | 3 | 2 |
| **.0** | .6 | .0 | .0 | .0 | .2 | 1.0 | .0 | .0 | hipB | 2 | 1 | 1 |
| **.0** | .4 | .0 | .0 | .0 | .8 | 1.0 | .0 | .1 | lrp | 2 | 2 | 1 |
| **.0** | .2 | .0 | .0 | .0 | .5 | 1.0 | .0 | .0 | ompR | 2 | 1 | 1 |
| **.0** | .2 | .0 | .0 | .0 | .7 | 1.0 | .0 | .0 | fur | 2 | 1 | 1 |
| **.2** | .3 | .0 | .0 | .0 | .7 | 1.0 | .0 | .0 | arcA | 2 | 1 | 1 |
| **.3** | .3 | .0 | .0 | .0 | .6 | 1.0 | .0 | .0 | phoB | 2 | 1 | 1 |
| **.1** | .4 | .0 | .0 | .0 | .6 | 1.0 | .0 | .1 | tyrR | 2 | 1 | 1 |
| **.3** | .5 | .2 | .2 | .2 | .8 | 1.0 | .1 | .4 | rpoD16 | 2 | 2 | 1 |
| **.3** | .5 | .1 | .1 | .1 | .8 | 1.0 | .1 | .3 | argR | 2 | 2 | 1 |
| **.0** | .3 | .2 | .2 | .2 | .8 | 1.0 | .2 | .4 | argR2 | 2 | 2 | 1 |
| **.3** | .3 | .2 | .1 | .2 | 1.0 | 1.0 | .1 | .3 | lexA | 2 | 2 | 2 |

**Table 5.** Ten different **a**ttribute weighting models applied on five samples showing the most important transcription factors based on various levels of weights assigned by each model.

| PCA | SVM | Relief | Uncertainty | Gini Index | Chi Squared | Deviation | Rule | Info Gain Ratio | Info Gain | Attribute | Count 50% | Count 75 | Count 95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **.8** | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | .8 | 1.0 | 1.0 | 1.0 | ihf | 10 | 10 | 8 |
| **.2** | 1.0 | .1 | .1 | .0 | .1 | .5 | .0 | .2 | .1 | arcA | 1 | 1 | 1 |
| **.3** | 1.0 | .1 | .3 | .2 | .3 | .6 | .4 | .3 | .2 | rpoD17 | 2 | 1 | 1 |
| **.5** | .6 | .1 | .4 | .4 | .4 | 1.0 | .6 | .4 | .4 | lrp | 4 | 1 | 1 |
| **.6** | .1 | .0 | .3 | .1 | .4 | 1.0 | .6 | .2 | .1 | argR | 3 | 1 | 1 |
| **.7** | .0 | .0 | .3 | .1 | .4 | 1.0 | .6 | .2 | .1 | rpoD16 | 3 | 1 | 1 |
| **1.0** | .8 | .6 | .6 | .6 | .7 | .7 | .8 | .6 | .6 | argR2 | 10 | 3 | 1 |
| **1.0** | .8 | .6 | .6 | .6 | .7 | .7 | .8 | .6 | .6 | rpoD18 | 10 | 3 | 1 |
| **1.0** | .8 | .6 | .6 | .6 | .7 | .7 | .8 | .6 | .6 | crp | 10 | 3 | 1 |
| **1.0** | .8 | .6 | .6 | .6 | .7 | .7 | .8 | .6 | .6 | flHCD | 10 | 3 | 1 |

**Table 6.** Ten different attribute weighting models applied on five samples (data transformed into binominal – YES/NO) showing the most important transcription factors based on various levels of weights assigned by each model.

| Relief | Uncertainty | Gini Index | Chi Squared | Rule | Info Gain Ratio | Info Gain | Attribute | Count 50% | Count 75 | Count 95 |
|---|---|---|---|---|---|---|---|---|---|---|
| **.5** | .6 | .7 | .7 | .1 | .6 | .6 | argR2 | 6 | 0 | 0 |
| **.5** | .6 | .7 | .7 | .1 | .6 | .6 | rpoD18 | 6 | 0 | 0 |
| **.5** | .6 | .7 | .7 | .1 | .6 | .6 | crp | 6 | 0 | 0 |
| **.5** | .6 | .7 | .7 | .1 | .6 | .6 | flHCD | 6 | 0 | 0 |
| **1.0** | 1.0 | 1.0 | 1.0 | .0 | 1.0 | 1.0 | ihf | 6 | 6 | 6 |

**PD200Genes**

The best performances in predicting the right genes obtained when Decision Tree model applied on PD200Genes dataset. As seen in Table 7, the accuracy of this model on predicting surfactin gene

was around 80% while the same accuracy for predicting other genes reached at 99.5%. As seen in Fig 4., Decision Tree model drew an inverted tree with *flHCD* feature at the tree root; showing when this feature was less to or equal to 0.5, the class was other genes but when it was higher than 0.5, if *rpoD15* was higher than 0.5, the class was surfacin gene, otherwise other genes.

**Table 7.** Confusion matrix for PD200Genes showing the accuracy of model in predicting the right class of other genes and surfactin genes (accuracy: 99.00% +/- 2.00% (mikro: 99.02%))

| | True other genes | True surfactin gene | Class precision |
|---|---|---|---|
| **Pred. other genes** | 198 | 1 | 99.50% |
| **Pred. surfactin gene** | 1 | 4 | 80.00% |
| **Class recall** | 99.50% | 80.00% | |

## PD5NGenes

The accuracy of Random forest model in predicting the right gene class for this dataset was 90%; with the best possible accuracy for predicting surfactin genes and 83.33% accuracy for predicting the other class (other genes) (Table 8).

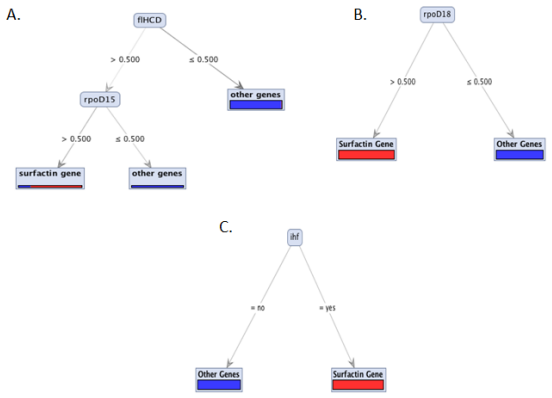



**Figure 4**. Decision trees induced by tree generating algorithms; showing the most important features in separation of surfactin gene. A. A decision tree that did not separate two classes of genes (surfactin and other genes) completely B. A decision tree separated two classes of genes (surfactin and other genes) completely. C. A decision tree confirmed the results of weighting algorithms

The model induced a simple one level tree with the *rpoD18* feature at the top. When this model was higher than 0.5, the gene class was other genes but when it was higher than 0.5, the gene class was surfactin genes.

## PD5BGenes

The best accuracy obtained when Decision Stump algorithm applied on this dataset, the model accuracy reached 90%; the model was perfect in predicting surfactin genes but less accurate in predicting the other genes (83.33%) (Table 8).
Decision Stump algorithm generated a simple tree with just ihf attribute at the root; this tree was capable of predicting the right class with just this feature (ihf), when it was yes, the gene class was surfactin, otherwise other genes (Fig 4).

**Table 8.** Confusion matrix for PD5NGenes and PD5BGene showing the accuracy of model in predicting the right class of other genes and surfactin genes (accuracy: 90.00% +/- 30.00% (mikro: 90.00%)

| | true Other Genes | true Surfactin Gene | class precision |
|---|---|---|---|
| **Pred. other genes** | 5 | 1 | 83.33% |
| **Pred. surfactin gene** | 0 | 4 | 100.00% |
| **Class recall** | 100.00% | 80.00% | |

## Discussion

Bacteria make a wide range of surface active compounds called "biosurfactants ". *Bacillus subtilis* produced surfactin as one of the most popular biosurfactant (Shao et al., 2015). Surfactin can reduce the surface and interfacial tension. Moreover, because of its tremendous potential, it is of great industrial and commercial interest (Yeh et al., 2005). Hence, molecular characterization of surfactin production regulatory elements and network is in demand.
Recently, we have isolated the *Bacillus subtilis* subsp. MJ01 from crude oil contaminated soil in the south of Iran (data not published). In previous study, we sequenced this strain with PacBio RS (Pacific Biosciences) and also annotating of this strain was performed too (data not published). The genome NCBI genbank accession number is CP018173. Laboratory experiments revealed the high amount of surfactin production and also the high Critical Micelle Concentration (CMC) of produced surfactin in distilled water for *B. subtilis subsp. MJ01*. In this study, we focused on promoter analysis and generate

the regulatory gene network for surfactin production for the first time to further investigate about molecular mechanisms of surfactin production. We expected to find some gene sequence alterations in surfactin production key genes between *B. subtilis* subsp. MJ01, spizizenii and 168, in order to explain the difference of surfactin production. Analysis of gene sequences such as *srfA*, *srfAA*, *srfAB*, *srfAD* and *sfp* involved in surfactin production could not reveal any significant variation between MJ01and spizizenii. However, about 7% difference was observed in these gene sequences in MJ01 and spizizenii compare with 168. *comQ* and *comX* were two genes which showed difference in both nucleotide and protein sequences between MJ01 and spizizenii vs 168. In addition, the only difference between MJ01 and spizizenii was observed in protein sequence of *comQ* and *comX*. These genes play the key roles for both competence and surfactin production (Oslizlo et al., 2014; Weinrauch et al., 1991). The findings of previous studies have validated the role of *comQ* and *comX* in surfactin production and our analysis identified them as a hot gene, strongly support the reliability of our results.

It has been suggested that TFs and promoter activating pattern may alter instead of alteration in coding sequences to generate more virulent strain (Mahdi et al., 2014; van Schaik et al., 2007). Hence, the identified potential TFBs and their organization open a new avenue to understand gene expression and regulation during surfactin production. By functional genomics based approach, we could compare TFs activating pattern of *Bacillus subtilis* MJ01 with spizizenii and 168, during surfactin production.

The most amazing result was not observed by difference in gene sequences but in the promoter and TFs patterns of key genes in surfactin production. The number of TFs activated on promoters of key genes involved in surfactin production (*srfA*, *srfAA*, *srfAB* and *srfAD*) was similar for each strain. It is obvious as all of these genes are the most important genes for surfactin production. But the number of TFs promote these genes (*srfA*, *srfAA*, *srfAB* and *srfAD*) was higher in MJ01 and spizizenii compare with 168. Mahdi et al. 2014 suggest that the higher number of TFs can be an index for more active/key genes (Mahdi et al., 2014).

Our findings also revealed that *Lrp*, *argR*, *rpoD*, *purR* and *ihf* were highly activated in all 3 strains. We suggest that these TFs might play a key role in surfactin production pathway based on their numbers. *rpoD* has been identified as the most active TFs among all 3 strains. It promotes transcription by attachment of RNA polymerase to the specific initiation site (Hengge-Aronis, 2002). *Ihf* wraps DNA around the body of the protein to form a higher-order nucleoprotein complex (Pagel et al., 1992; Winkelman and Hatfield, 1990) and facilitates the unwinding of the DNA helix in the -10 hexanucleotide region of the downstream promoter (Parekh and Hatfield, 1996). Also Parekh et al. 1996 suggest that Ihf activates transcription of some genes by forming a higher order protein-DNA complex that change the DNA helix in order to assist opening DNA helix at downstream promoters site (Parekh et al., 1996). *Lrp* has role in global regulation of cellular metabolism (Calvo and Matthews, 1994).

The observed difference between the predicted TF activation patterns among 3 strains suggest that different strains of *Bacillus subtilis* may activate different pattern of regulatory element for producing their surfactin.

However, the pattern of TFs for surfactin producing genes showed completely different among these 3 strains (MJ01, spizizenii and 168). Therefor we can conclude that the both number and pattern of TFs might be important for regulation of surfactin process.

*lexA* has role in transferring of mobile genetic elements and also involve in formation of biofilm. Moreover, it represses the number of genes involved in response to DNA damage (SOS response) (Mo et al., 2014). Also *rpoD* involves in promoting of RNA polymerase to attach to specific initiation site. In addition, it play a key role in transcription of growth related genes(Shimada et al., 2014). The network analysis revealed that these TFs are hub. Their general role in transcription approves our results.

In addition data mining and decision tree revealed that *flHCD*, *rpoD* and *ihf* are the most important TF in order to distinct surfactin producing gene and other genes in *B.subtilis* subsp. MJ01. Promoter analysis in 3 *B. subtilis* subsp. 168, spizizenii and MJ01 also identified these three TFs (*flHCD*, *rpoD* and *ihf*) as the most important and abundant TF among all three strains.

Hence in this research, the result of both data mining and decision tree among general genes and key surfactin producing genes confirmed the result of TF patterns analysis of key surfactin producing genes between spizizenii, 168 and MJ01. However, all analysis identified *flHCD*, *rpoD* and *ihf* as the most important TF for surfactin producing genes.

Also network analysis for surfactin producing genes occurred. Two modules have been constructed. Module 1 and 2 covered 44% and 28% of global network. Module 1 includes surfactin production key genes (*srfA*, *srfAA*, *srfAB*, *srfAD* and *sfp*) with key

TFs that identified by our promoter analysis such as *rpoD*, *ihf*, *arcA*, *cytR*, *rpoH* and *argR2*. In addition, *ppsA* and *ppsC* which are activated in module 1, encode the nonribosomal peptide synthetase (NRPS) subunits (Du and Shen, 2001). However, module 2 consists of just *srfA* and some regulatory genes relate to producing surfactin such as *comQ* and *comX*.
The isoprenyl transferase *ComQ* modifies the signaling peptide *ComX*. The isoprenylated *ComX* is then secreted (Magnuson et al., 1994) and by the time the concentration reached at critical point the auto-phosphorylation of the membrane-bound *ComP* would be activated, that can phosphorylate the transcriptional activator *ComA* (Weinrauch et al., 1990). Phosphorylated *ComA* directly regulates the expression of various genes, such as the *srfA* operon (Oslizlo et al., 2014). Moreover, *rpoD*, *fis*, *purr*, *argR*, *lexA* and *lrp* are the available TFs in module 2 that our analysis confirmed their role in surfactin production.

## Conclusion

In this study, promoter and network analysis, opened for the first time a new avenue for understanding the molecular mechanism of surfactin production in *Bacillus subtilis*. Specially, comparison between 3 strains of *Bacillus subtilis* (MJ01, spizizenii and 168), revealed that *Bacillus subtilis* subsp. MJ01 which we could isolate from south oil contaminated soil of Iran, have potential to be the novel strain, although it requires more studies.

## References

1. Abdel-Mawgoud A. M., Aboulwafa M. M. and Hassouna N. A.-H. (2008) Characterization of surfactin produced by *Bacillus subtilis* isolate BS5. Applied Biochemistry and Biotechnology 150:289-303.
2. Anuradha S N. (2010) Structural and molecular characteristics of lichenysin and its relationship with surface activity. Biosurfactants:304-315.
3. Banat I. M., Franzetti A., Gandolfi I., Bestetti G., Martinotti M. G., Fracchia L., Smyth T. J. and Marchant R. (2010) Microbial biosurfactants production, applications and future potential. Applied Microbiology and Biotechnology 87:427-444.
4. Calvo J. M. and Matthews R. G. (1994) The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiological Reviews 58:466-490.
5. Cao X.-h., Wang A.-h., Wang C.-l., Mao D.-z., Lu M.-f., Cui Y.-q. and Jiao R.-z. (2010) Surfactin induces apoptosis in human breast cancer MCF-7 cells through a ROS/JNK-mediated mitochondrial/caspase pathway. Chemico-Biological Interactions 183:357-362.
6. Deihimi T., Niazi A., Ebrahimi M., Kajbaf K., Fanaee S., Bakhtiarizadeh M. R. and Ebrahimie E. (2012) Finding the undiscovered roles of genes: an approach using mutual ranking of coexpressed genes and promoter architecture-case study: dual roles of thaumatin like proteins in biotic and abiotic stresses. SpringerPlus 1:1.
7. Du L. and Shen B. (2001) Biosynthesis of hybrid peptide-polyketide natural products. Current Opinion in Drug Discovery & Development 4:215-228.
8. Ebrahimi M., Ebrahimie E. and Bull C. M. (2015) Minimizing the cost of translocation failure with decision-tree models that predict species' behavioral response in translocation sites. Conservation Biology 29:1208-1216.
9. Hengge-Aronis R. (2002) Signal transduction and regulatory mechanisms involved in control of the σS (RpoS) subunit of RNA polymerase. Microbiology and Molecular Biology Reviews 66:373-395.
10. Lee W. and Chen S. L. (2002) Research report genome-tools: A flexible package for genome sequence analysis. Biotechniques 33:1334-1341.
11. Magnuson R., Solomon J. and Grossman A. D. (1994) Biochemical and genetic characterization of a competence pheromone from B. subtilis. Cell 77:207-216.
12. Mahdi L. K., Deihimi T., Zamansani F., Fruzangohar M., Adelson D. L., Paton J. C., Ogunniyi A. D. and Ebrahimie E. (2014) A functional genomics catalogue of activated transcription factors during pathogenesis of pneumococcal disease. BMC Genomics 15:1.
13. Mo C. Y., Birdwell L. D. and Kohli R. M. (2014) Specificity determinants for autoproteolysis of LexA, a key regulator of bacterial SOS mutagenesis. Biochemistry 53:3158-3168.
14. Mulligan C. N. (2009) Recent advances in the environmental applications of biosurfactants. Current Opinion in Colloid & Interface Science 14:372-378.
15. Nepusz T., Yu H. and Paccanaro A. (2012) Detecting overlapping protein complexes in protein-protein interaction networks. Nature Methods 9:471-472.
16. Nitschke M. and Costa S. (2007) Biosurfactants in food industry. Trends in Food Science & Technology 18:252-259.
17. Oslizlo A., Stefanic P., Dogsa I. and Mandic-Mulec I. (2014) Private link between signal and response in *Bacillus subtilis* quorum sensing. Proceedings of the National Academy of Sciences 111:1586-1591.
18. Pagel J. M., Winkelman J. W., Adams C. W. and Hatfield G. W. (1992) DNA topology-mediated regulation of transcription initiation from the tandem promoters of the ilvGMEDA operon of *Escherichia coli*. Journal of Molecular Biology 224:919-935.

19. Parekh B. S. and Hatfield G. W. (1996) Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model. Proceedings of the National Academy of Sciences 93:1173-1177.
20. Parekh B. S., Sheridan S. D. and Hatfield G. W. (1996) Effects of integration host factor and DNA supercoiling on transcription from the ilvPG promoter of *Escherichia coli*. Journal of Biological Chemistry 271:20258-20264.
21. Pashaiasl M., Ebrahimi M. and Ebrahimie E. (2016a) Identification of the key regulating genes of diminished ovarian reserve (DOR) by network and gene ontology analysis. Molecular Biology Reports 43:923-937.
22. Pashaiasl M., Khodadadi K., Kayvanjoo A. H., Pashaei-asl R., Ebrahimie E. and Ebrahimi M. (2016b) Unravelling evolution of Nanog, the key transcription factor involved in self-renewal of undifferentiated embryonic stem cells, by pattern recognition in nucleotide and tandem repeats characteristics. Gene 578:194-204.
23. Porob S., Nayak S., Fernandes A., Padmanabhan P., Patil B. A., Meena R. M. and Ramaiah N. (2013) PCR screening for the surfactin (sfp) gene in marine Bacillus strains and its molecular characterization from Bacillus tequilensis NIOS11. Turkish Journal of Biology 37:212-221.
24. Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J. T., Ramage D., Amin N., Schwikowski B. and Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13:2498-2504.
25. Shao C., Liu L., Gang H., Yang S. and Mu B. (2015) Structural diversity of the microbial surfactin derivatives from selective esterification approach. International Journal of Molecular Sciences 16:1855-1872.
26. Shibulal B., Al-Bahry S. N., Al-Wahaibi Y. M., Elshafie A. E., Al-Bemani A. S. and Joshi S. J. (2014) Microbial enhanced heavy oil recovery by the aid of inhabitant spore-forming bacteria: an insight review. The Scientific World Journal 2014.
27. Shimada T., Yamazaki Y., Tanaka K. and Ishihama A. (2014) The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of *Escherichia coli*. PLoS One 9:e90447.
28. Shoemaker B. A. and Panchenko A. R. (2007) Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLOS Computational Biology 3:e43.
29. Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., Simonovic M., Roth A., Santos A. and Tsafou K. P. (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research:gku1003.
30. Torkzaban B., Kayvanjoo A. H., Ardalan A., Mousavi S., Mariotti R., Baldoni L., Ebrahimie E., Ebrahimi M. and Hosseini-Mazinani M. (2015) Machine learning based classification of microsatellite variation: an effective approach for phylogeographic characterization of olive populations. PloS One 10:e0143465.
31. Van Schaik W., van der Voort M., Molenaar D., Moezelaar R., de Vos W. M. and Abee T. (2007) Identification of the σB regulon of Bacillus cereus and conservation of σB-regulated genes in low-GC-content gram-positive bacteria. Journal of Bacteriology 189:4384-4390.
32. Weinrauch Y., Msadek T., Kunst F. and Dubnau D. (1991) Sequence and properties of comQ, a new competence regulatory gene of *Bacillus subtilis*. Journal of Bacteriology 173:5685-5693.
33. Weinrauch Y., Penchev R., Dubnau E., Smith I. and Dubnau D. (1990) A *Bacillus subtilis* regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems. Genes & Development 4:860-872.
34. Winkelman J. and Hatfield G. W. (1990) Characterization of the integration host factor binding site in the ilvPG1 promoter region of the ilvGMEDA operon of *Escherichia coli*. Journal of Biological Chemistry 265:10055-10060.
35. Yada T., Totoki Y., Ishii T. and Nakai K. 1997. Functional prediction of *B. subtilis* genes from their regulatory sequences. *In* ISMB. Vol. 5. 354-357.
36. Yeh M. S., Wei Y. H. and Chang J. S. (2005) Enhanced Production of Surfactin from Bacillussubtilis by addition of solid carriers. Biotechnology Progress 21:1329-1334.

**Supplementary 1**. %sequence identity of genes associate with key surfactin producing genes in their networks between 168 vs spizizenii and MJ01

| Gene name | Gene definition | Blastn result (% Identity) | |
|---|---|---|---|
| | | Spizizenii | MJ01 |
| *pksN* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 94.38 | 93.90 |
| *pksM* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 93 | 93.03 |
| *pksJ* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 96.68 | 96.27 |
| *pksD* | polyketide synthase (EC:2.3.1.-); Probably involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 94.20 | 94.12 |
| *pksL* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 92.95 | 92.95 |
| *pksR* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 90.79 | 90.92 |
| *ppsC* | plipastatin synthetase; This protein is a multifunctional enzyme, able to activate and polymerize the amino acids Glu and Ala/Val as part of the biosynthesis of the lipopeptide antibiotic plipastatin. The Ala/Val residue is further epimerized to the D-isomer form. The activation sites for these amino acids consist of individual domains | 89.04 | 89.18 |
| *ppsA* | plipastatin synthetase; This protein is a multifunctional enzyme, able to activate and polymerize the amino acids Glu and Orn as part of the biosynthesis of the lipopeptide antibiotic lipastatin. The Orn residue is further epimerized to the D-isomer form. The activation sites for these amino acids consist of individual domains | 89.27 | 89.33 |
| *comQ* | isoprenyl transferase; Involved in the maturation of ComX, part of a major quorum-sensing system that regulates the development of genetic competence | 81.17 | 80.86 |
| *comX* | competence pheromone precursor (pheromone peptide aa 46->55, modified); Part of a major quorum-sensing system that regulates the development of genetic competence. Acts through the activation of the two-component regulatory system ComP/ComA composed of a sensor histidine kinase, ComP, and a response regulator, ComA, that regulates directly the transcription of over 20 genes. Transport through the membrane may involve Spo0K. Under certain conditions plays a role in sporulation | 87.31 | 87.30 |
| *yndJ* | hypothetical protein | 88.26 | 88.49 |
| *ybdz* | hypothetical protein | 96.19 | 96.19 |
| *menB* | naphthoate synthase (EC:4.1.3.36); Converts o-succinylbenzoyl-CoA (OSB-CoA) to 1,4- dihydroxy-2-naphthoyl-CoA (DHNA-CoA) | 93.53 | 93.41 |
| *dltA* | D-alanine--poly(phosphoribitol) ligase subunit 1 (EC:6.1.1.13); Involved in the biosynthesis of D- | 93.77 | 93.85 |

| | | | |
|---|---|---|---|
| | alanyl-lipoteichoic acid (LTA). Catalyzes an ATP-dependent two-step reaction where it forms a high energy D-alanyl AMP intermediate and transfers the alanyl residues from AMP to Dcp | | |
| *dhbB* | isochorismatase (EC:3.3.2.1) | 92.76 | 93.29 |
| *codY* | transcriptional repressor CodY; DNA-binding protein that represses the expression of many genes that are induced as cells make the transition from rapid exponential growth to stationary phase and sporulation. It is a GTP-binding protein that senses the intracellular GTP concentration as an indicator of nutritional limitations. At low GTP concentration it no longer binds GTP and stop to act as a transcriptional repressor | | |
| *RapC* | response regulator aspartate phosphatase | 95.13 | 95.13 |
| *dhbF* | siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin synthetase (EC:2.7.7.-5.1.1.-); Specifically adenylates threonine and glycine, and loads them onto their corresponding peptidyl carrier domains | 91.89 | 91.93 |
| *pksS* | cytochrome P450; Involved in the metabolism of the antibiotic polyketide bacillaene which is involved in secondary metabolism. The substrate is dihydrobacillaene | 92.53 | 92.69 |
| *mecA* | adaptor protein; Enables the recognition and targeting of unfolded and aggregated proteins to the ClpC protease or to other proteins involved in proteolysis. Acts negatively in the development of competence by binding ComK and recruiting it to the ClpCP protease. When overexpressed, inhibits sporulation. Also involved in Spx degradation by ClpC | 98.48 | 98.78 |
| *comK* | competence transcription factor (CTF); Intermediate regulatory gene required for the expression of the late competence genes comC, comE, comG and the bdbDC operon. Receives signals from SrfA, and possibly other regulatory COM genes, and transduces these signals to the late COM genes | 94.82 | 94.99 |
| *YCZE* | integral inner membrane protein regulating antibiotic production | 95.83 | 95.68 |

**Supplementary 2**. %identity of protein sequence of genes associate with key surfactin producing genes in their networks between 168 vs spizizenii and MJ01

| **Gene name** | **Gene definition** | **Tblastn result (% Identity)** | |
|---|---|---|---|
| | | **Spizizenii** | **MJ01** |
| *pksN* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 90.79 | 90.92 |
| *pksM* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 89.04 | 89.18 |
| *pksJ* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 89.27 | 89.33 |
| *pksD* | polyketide synthase (EC:2.3.1.-); Probably involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 81.17 | 80.86 |

| | | | |
|---|---|---|---|
| *pksL* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 87.31 | 87.30 |
| *pksR* | polyketide synthase; Involved in some intermediate steps for the synthesis of the antibiotic polyketide bacillaene which is involved in secondary metabolism | 88.26 | 88.49 |
| *ppsC* | plipastatin synthetase; This protein is a multifunctional enzyme, able to activate and polymerize the amino acids Glu and Ala/Val as part of the biosynthesis of the lipopeptide antibiotic plipastatin. The Ala/Val residue is further epimerized to the D-isomer form. The activation sites for these amino acids consist of individual domains | 38.98 | 38.91 |
| *ppsA* | plipastatin synthetase; This protein is a multifunctional enzyme, able to activate and polymerize the amino acids Glu and Orn as part of the biosynthesis of the lipopeptide antibiotic lipastatin. The Orn residue is further epimerized to the D-isomer form. The activation sites for these amino acids consist of individual domains | 38.68 | 38.69 |
| *comQ* | isoprenyl transferase; Involved in the maturation of ComX, part of a major quorum-sensing system that regulates the development of genetic competence | 45.59 | 91.30 |
| *comX* | competence pheromone precursor (pheromone peptide aa 46->55, modified); Part of a major quorum-sensing system that regulates the development of genetic competence. Acts through the activation of the two-component regulatory system ComP/ComA composed of a sensor histidine kinase, ComP, and a response regulator, ComA, that regulates directly the transcription of over 20 genes. Transport through the membrane may involve Spo0K. Under certain conditions plays a role in sporulation | 27.08 | 90.57 |
| *yndJ* | hypothetical protein | 89.78 | 90.15 |
| *ybdz* | hypothetical protein | 97 | 97 |
| *menB* | naphthoate synthase (EC:4.1.3.36); Converts o-succinylbenzoyl-CoA (OSB-CoA) to 1,4- dihydroxy-2-naphthoyl-CoA (DHNA-CoA) | 99.26 | 99.26 |
| *dltA* | D-alanine--poly(phosphoribitol) ligase subunit 1 (EC:6.1.1.13); Involved in the biosynthesis of D-alanyl-lipoteichoic acid (LTA). Catalyzes an ATP-dependent two-step reaction where it forms a high energy D-alanyl AMP intermediate and transfers the alanyl residues from AMP to Dcp | 96.42 | 96.42 |
| *dhbB* | isochorismatase (EC:3.3.2.1) | 94.55 | 94.87 |
| *codY* | transcriptional repressor CodY; DNA-binding protein that represses the expression of many genes that are induced as cells make the transition from rapid exponential growth to stationary phase and sporulation. It is a GTP-binding protein that senses the intracellular GTP concentration as an indicator of nutritional limitations. At low GTP concentration it no longer binds GTP and stop to act as a transcriptional repressor | 100 | 100 |
| *RapC* | response regulator aspartate phosphatase | 97.91 | 97.91 |
| *dhbF* | siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin synthetase (EC:2.7.7.-5.1.1.-); Specifically adenylates threonine and glycine, and loads them onto their corresponding peptidyl carrier domains | 94.07 | 93.95 |

| | | | |
|---|---|---|---|
| *pksS* | cytochrome P450; Involved in the metabolism of the antibiotic polyketide bacillaene which is involved in secondary metabolism. The substrate is dihydrobacillaene | 95.06 | 95.31 |
| *mecA* | adaptor protein; Enables the recognition and targeting of unfolded and aggregated proteins to the ClpC protease or to other proteins involved in proteolysis. Acts negatively in the development of competence by binding ComK and recruiting it to the ClpCP protease. When overexpressed, inhibits sporulation. Also involved in Spx degradation by ClpC | 99.08 | 99.08 |
| *comK* | competence transcription factor (CTF); Intermediate regulatory gene required for the expression of the late competence genes comC, comE, comG and the bdbDC operon. Receives signals from SrfA, and possibly other regulatory COM genes, and transduces these signals to the late COM genes | 99.08 | 99.08 |
| *YCZE* | integral inner membrane protein regulating antibiotic production | 95.81 | 95.81 |

# Scientific Reviewers

# MANUSCRIPT PREPARATION

Manuscripts should be prepared in accordance with the uniform requirements for Manuscript's Submission to **"Journal of Cell and Molecular Research".**

**Language:** Papers should be in English (either British or American spelling). The past tense should be used throughout the results description, and the present tense in referring to previously established and generally accepted results. Authors who are unsure of correct English usage should have their manuscript checked by somebody who is proficient in the language; manuscripts that are deficient in this respect may be returned to the author for revision before scientific review.

**Typing:** Manuscripts must be typewritten in a font size of at least 12 points, double-spaced (including References, Tables and Figure legends) with wide margins (2.5 cm from all sides) on one side of the paper. The beginning of each new paragraph must be clearly indicated by indentation. All pages should be numbered consecutively at the bottom starting with the title page.

**Length:** The length of research articles should be restricted to ten printed pages. Short communication should not exceed five pages of manuscript, including references, figures and tables. Letters should be 400-500 words having 7-10 references, one figure or table if necessary. Commentaries and news should also be 800-1000 words having 7-10 references and one figure or table if necessary.

**Types of Manuscript:** JCMR is accepting original research paper, short communication reports, invited reviews, letters to editor, biographies of scientific reviewers, commentaries and news.

**Statement of Human and Animal Rights**: Author's should declare regulatory statement regarding the experiments using animals, human cells/tissues that all in vivo experiments have been performed according to the guidelines (explained by WHO, international animal rights federations or your respective institute) to use animals in their research work.

**Conflict of Interest Statement**: Authors or corresponding author should declare statement of conflict of interest at the last of manuscript.

**Manuscript Evaluation Time**: All submitted manuscripts will be evaluated and reviewed according to following evaluation schedule.
Pre-Editorial Evaluation: All submitted manuscripts, right after their submission to JCMR will be evaluation by Editors for being according to the journal scope and format. This evaluation can take 2-7 days of submission.
Reviewer's Evaluation: Selected manuscripts after pre-editorial evaluation will be sent to minimum two blind reviewers assigned by Editor-in-Chief. This process may take 21-27 days.
Post Editorial Evaluation: After receiving reviewer's comments, editors evaluate the manuscripts considering the comments and decide their first decision. This process takes 3-5 days and then authors are informed regarding the editorial decision.

## GENERAL ARRANGEMENT OF PAPERS

**Title:** In the first page, papers should be headed by a concise and informative title. The title should be followed by the authors' full first names, middle initials and last names and by names and addresses of laboratories where the work was carried out. Identify the affiliations of all authors and their institutions, departments or organization by use of Arabic numbers (1, 2, 3, etc.).

**Footnotes:** The name and full postal address, telephone, fax and E-mail number of corresponding author should be provided in a footnote.

**Abbreviations:** The Journal publishes a standard abbreviation list at the front of every issue. These standard abbreviations do not need to be spelled out within paper. However, non-standard and undefined abbreviations used five or more times should be listed in the footnote. Abbreviations should be defined where first mentioned in the text. Do not use abbreviations in the title or in the Abstract. However, they can be used in Figures and Tables with explanation in the Figure legend or in a footnote to the Table.

**Abstract:** In second page, abstract should follow the title (no authors' name) in structured format of not more than 250 words and must be able to stand independently and should state the Background, Methods, Results and Conclusion. Write the abstract in third person. References should not be cited and abbreviations should be avoided.

**Keywords:** A list of three to five keywords for indexing should be included at bottom of the abstract. Introduction should contain a description of the problem under investigation and a brief survey of the existing literature on the subject.

**Materials and Methods:** Sufficient details must be provided to allow the work to be repeated. Correct chemical names should be given and strains of organisms should be specified. Suppliers of materials need only be mentioned if this may affect the results. Use System International (SI) units and symbols.

**Results:** This section should describe concisely the rationale of the investigation and its outcomes. Data should not be repeated in both a Table and a Figure. Tables and Figures should be selected to illustrate specific points. Do not tabulate or illustrate points that can be adequately and concisely described in the text.

**Discussion:** This should not simply recapitulate the Results. It should relate results to previous work and interpret them. Combined Results and Discussion sections are encouraged when appropriate.

**Acknowledgments:** This optional part should include a statement thanking those who assisted substantially with work relevant to the study. Grant support should be included in this section.

**References:** References should be numbered and written in alphabetical order. Only published, "in press" papers, and books may be cited in the reference list (see the examples below). References to work "in press" must be accompanied by a copy of acceptance letter from the journal. References should not be given to personal communications, unpublished data, manuscripts in preparation, letters, company publications, patents pending, and URLs for websites. Abstracts of papers presented at meetings are not permissible. These references should appear as parenthetical expressions in the text, e.g. (unpublished data). Few example of referencing patterns are given as follows:

Bongso A., Lee E. H. and Brenner S. (2005) Stem cells from bench to bed side. World Scientific Publishing Co. Singapore, 38-55 pp.

Irfan-Maqsood M. (2013) Stem Cells of Epidermis: A Critical Introduction. Journal of Cell and Molecular Research 5(1): 1-2.

**Note:** All the reference should be in EndNote format (JCMR EndNote Style is available on JCMR's web site, Author's Guideline)

**Tables and Figures:** Tables and Figures should be numbered (1, 2, 3, etc.) as they appear in the text. Figures

should preferably be the size intended for publication. Tables and Figures should be carefully marked. Legends should be typed single-spaced separately from the figures. Photographs must be originals of high quality. Photocopies are not acceptable. Those wishing to submit color photographs should contact the Editor regarding charges.

**Black Page Charges:** There is no black page charges for publication in the Journal of Cell and Molecular Research.

**Color Page Charges:** All color pages being printed in color will cost 1,000,000 Iranian Rials/page.

**JCMR Open Access Policy**: Journal of Cell and Molecular Research follows the terms outlined by the Creative Common's Attribution-Only license (CC-BY) to be the standard terms for Open Access. Creative Commons License.



Note: All manuscripts submitted to JCMR are tracked by using "Plagiarism Tracker X" for possible plagiarism before acceptance to JCMR

# Table of Contents



Journal of Cell and Molecular Research

Volume 9 , Number 2, Winter 2017