J. Microbiol. Biotechnol. (2015), 25(10), 1599–1605
http://dx.doi.org/10.4014/jmb.1504.04008

jmb

Comparative Genomics Reveals the Core and Accessory Genomes of *Streptomyces* Species S

Ji-Nu Kim[1], Yeonbum Kim[2], Yujin Jeong[3], Jung-Hye Roe[2], Byung-Gee Kim[1], and Byung-Kwan Cho[3]*

[1]*School of Chemical and Biological Engineering, Institute of Molecular Biology and Genetics, and Bioengineering Institute, Seoul National University, Seoul 151-742, Republic of Korea*
[2]*Laboratory of Molecular Microbiology, School of Biological Sciences, Institute of Microbiology, Seoul National University, Seoul 151-742, Republic of Korea*
[3]*Department of Biological Sciences and KI for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea*

Received: April 6, 2015
Revised: May 27, 2015
Accepted: May 28, 2015

First published online
June 2, 2015

**Corresponding author*
Phone: +82-42-350-2620;
Fax: +82-42-350-5620;
E-mail: bcho@kaist.ac.kr

Supplementary data for this paper are available on-line only at http://jmb.or.kr.

pISSN 1017-7825, eISSN 1738-8872

The development of rapid and efficient genome sequencing methods has enabled us to study the evolutionary background of bacterial genetic information. Here, we present comparative genomic analysis of 17 *Streptomyces* species, for which the genome has been completely sequenced, using the pan-genome approach. The analysis revealed that 34,592 ortholog clusters constituted the pan-genome of these *Streptomyces* species, including 2,018 in the core genome, 11,743 in the dispensable genome, and 20,831 in the unique genome. The core genome was converged to a smaller number of genes than reported previously, with 3,096 gene families. Functional enrichment analysis showed that genes involved in transcription were most abundant in the *Streptomyces* pan-genome. Finally, we investigated core genes for the sigma factors, mycothiol biosynthesis pathway, and secondary metabolism pathways; our data showed that many genes involved in stress response and morphological differentiation were commonly expressed in *Streptomyces* species. Elucidation of the core genome offers a basis for understanding the functional evolution of *Streptomyces* species and provides insights into target selection for the construction of industrial strains.

Keywords: *Streptomyces*, comparative genomics, pan-genome, core genome

Introduction

Streptomycetes are active producers of a wide range of secondary metabolites, including more than two-thirds of the natural antibiotics in the pharmaceutical industry [1, 19]. They are members of the largest genus of actinobacteria, which are ubiquitous in soil and undergo complex differentiation from filamentous mycelia to aerial hyphae, and spores [6, 9]. For the genome-scale elucidation of the genetic background of secondary metabolites and the rich repertoire of novel enzymes in this genus, extensive sequence analyses have been carried out for different model *Streptomyces* strains, such as *Streptomyces coelicolor* A3(2) [1], *S. griseus* [24], and *S. avermitilis* [14]. In addition to a high G+C ratio and linear chromosome shape as important genomic features, the *Streptomyces* genome encodes a number of sigma factors and transcription factors that are involved in the complex transcriptional regulatory network [2]. Many genes are involved in morphological differentiation, and tens of gene clusters encode genes that participate in the biosynthesis of secondary metabolites in each strain [22].

To date, the genome sequences of over 30,000 bacterial species have been reported in the NCBI genome database (http://www.ncbi.nlm.nih.gov/genome/browse). From this abundance of information, comparative genomics analyses between multiple genomes of individual species have been used to reveal extensive genomic inter- and intraspecies diversity [3]. Among currently available comparative analysis methods, pan-genome analysis has

been used to describe the entire gene repertoire of bacterial species through identifying the sum of the core and dispensable genomes [21]. Thus, a pan-genome is defined as the full set of non-orthologous genes present in species, composed of the core and dispensable genomes; that is sets of genes that are present in all strains and unique to single strains, respectively. This analysis demonstrated how many new genes can be identified from newly sequenced genomes. Several reports of comparative genomic studies have revealed a catalog of genomic components and the evolutionary history of *Streptomyces* species [13, 15, 32]. However, even the most recent study analyzed five model *Streptomyces* spp. [32], requiring incorporation of current sequence information.

In this study, based on the rapidly increasing number of genomes sequenced, we performed comprehensive analysis of the genomes of all 17 *Streptomyces* species that have been completely sequenced to date in order to understand their genomic components. We estimated the pan-genome of *Streptomyces* and identified the core genome that was conserved in all of the analyzed strains. In addition, ortholog clusters within the pan-genome were classified according to their functions, and genes that showed distinctive characteristics of *Streptomyces* were listed. This analysis provides up-to-date information on genomic diversity and core conservation of *Streptomyces* genomes, facilitating our comprehensive understanding of this genus.

Materials and Methods

Nucleotide Sequence Accession Numbers

All the complete genome sequences of the 17 *Streptomyces* species used for our analysis were retrieved from NCBI FTP (ftp://ftp.ncbi.nih.gov/genome/Bacteria). The accession numbers for these 17 *Streptomyces* species are NC_021055 (*Streptomyces* sp. PAMC26508), NC_015953 (*Streptomyces* sp. SirexAA E), NC_020990 (*S. albus* J1074), NC_003155 (*S. avermitilis* MA 4680), NC_016582 (*S. bingchenggensis* BCW1), NC_016111 (*S. cattleya* NRRL 8057), NC_003888 (*S. coelicolor* A3(2)), NC_021985 (*S. collinus* Tü 365), NC_020504 (*S. davawensis* JCM 4913), NC_016114 (*S. flavogriseus* ATCC 33331), NC_021177 (*S. fulvissimus* DSM 40593), NC_010572 (*S. griseus* NBRC 13350), NC_017765 (*S. hygroscopicus jinggangensis* 5008), NC_022785 (*S. rapamycinicus* NRRL 5491), NC_013929 (*S. scabiei* 87 22), NC_018750 (*S. venezuelae* ATCC 10712), and NC_015957 (*S. violaceusniger* Tü 4113).

Pan-Genome Calculation

For the pan-genome computation of *Streptomyces* species, PGAP ver. 1.12 was used [31]. Ortholog clusters were organized using the open reading frame (ORF) contents of each genome with the GF (Gene Family) method using default parameters (E-value: 1e-10, score: 40; identity: 50; coverage: 50). The pan-genome and core genome profiles were then built. Functional enrichment of ortholog clusters was performed using the PGAP program and was used for the classification of clusters of orthologous groups (COGs). Subsequent classification work was performed using an in-house script.

Results and Discussion

The Pan-Genome of 17 *Streptomyces* Species

Seventeen completely sequenced *Streptomyces* species genomes available at the NCBI FTP database (ftp://ftp.ncbi.nih.gov/genomes/Bacteria) were used in this study. The genomic characteristics of each species are summarized in Supplementary Table S1. All strains are reported to contain linear chromosomes. Their genome sizes range from 6.3 to 12.7 Mb with G+C contents from 70.6% to 73.3%. The number of predicted coding sequences (CDSs; 5,832–10,022) was positively correlated with their genome size.

Pan-genome analysis of the 17 *Streptomyces* chromosomes revealed 34,592 ortholog clusters from 1,129,413 total genes that constituted the pan-genome. The size of the *Streptomyces* pan-genome may grow with the number of sequenced strains, and this pan-genome can therefore be considered an open pan-genome (Fig. 1A) [21]. This trend suggests that *Streptomyces* has flexible genome contents, reflecting the diversity of secondary metabolism and morphological differentiation, which is pronounced in this genus. The core genome consisted of 2,018 ortholog clusters (Fig. 1B and Table S2). This number is smaller than that in a previous report, which described 3,096 gene families based on five *Streptomyces* strains [32]. The ratio of the core genome in each species ranged from 24% to 38% and was negatively correlated with the number of ORFs. Although this number may be decreased when the analyzed genome is added, the number of core genomes would be expected to converge to a constant value, as judged from the slope of exponential decay. The number of dispensable gene families that were conserved in at least two species was 11,743, and the number of ortholog clusters of unique genes that were present in only one strain was 20,831. We called these two groups accessory genomes; these genomes are thought to contribute to the species' diversity and generally provide functions that were not essential to viability. However, these genes may have conferred a selective advantage to Streptomycetes in their specific environmental niche.

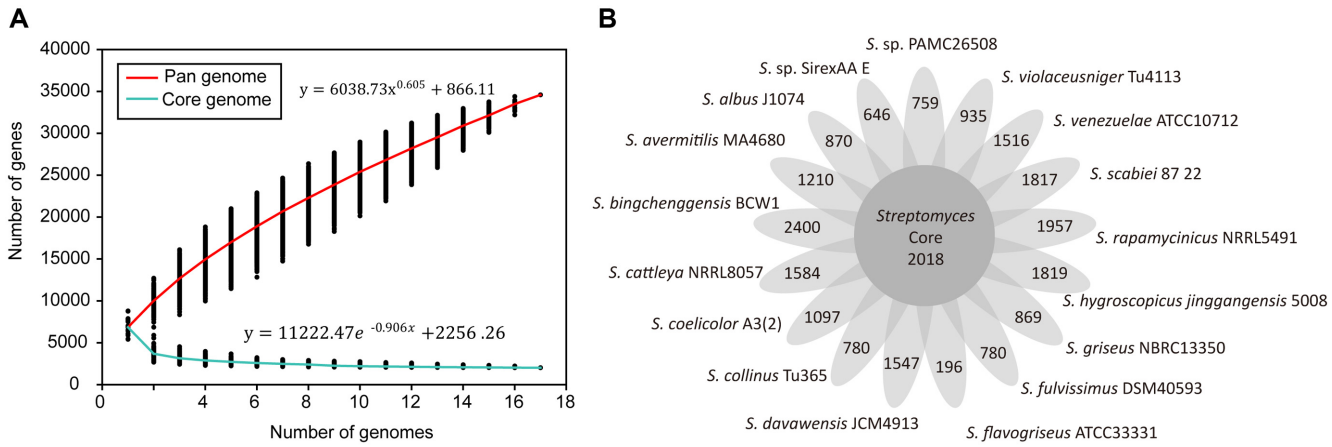

Fig. 1. Pan-genome analysis of *Streptomyces*.
(**A**) Pan-genome and core genome profiles. The numbers of new genes in the *Streptomyces* pan-genome and core genome are plotted against the number of genomes added. The deduced mathematical function is also reported. (**B**) Venn diagram showing the number of species-specific gene families in the genome of each species. The number of core genomes is represented in the center.

Functional Distribution of Ortholog Clusters

Next, we examined the functional classifications of ortholog clusters using the COG database (Table S3). The most abundant COG category in the pan-genome, except poorly or uncharacterized ones, was transcription (K) that included 1,945 gene families. The next abundant COGs were transport and metabolism of carbohydrates (G; 1,242) and amino acids (E; 1,046). In the core genome, the transcription category still encompassed the largest gene families (211), followed by metabolism of amino acids (192) and carbohydrates (130). The abundance of transcriptional regulators, including sigma factors, is a hallmark of Streptomycetes, consistent with their complex transcriptional regulatory networks that support morphological and physiological differentiation [2].

We then investigated the proportion of each conserved

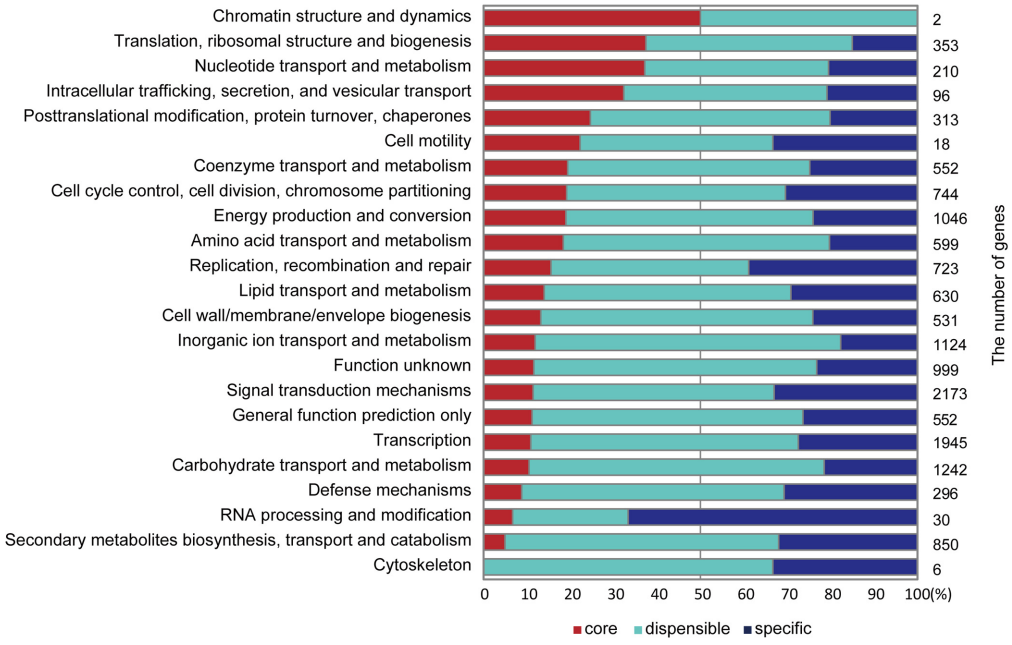

Fig. 2. Distribution of orthologous genes based on COG category.
The bars are sorted by the proportion of core genomes in each functional category.

group (core, dispensable, and unique genomes) to determine the numbers of genes in each category (Fig. 2). We found that the occurrence ratio in core genomes was high for the COG categories of translation (J) and nucleotide metabolism (F). This reveals the importance of protein and nucleic acid synthesis as the conserved core function, and the relatively lower diversity of genes within these categories. In comparison, those for secondary metabolism, defense mechanisms, carbohydrate transport/metabolism, and transcription occurred less frequently in the core genomes. Even though the absolute number of gene families in these categories is large, the finding that most of these genes reside in accessory genomes suggests that they provide functions to increase the diversity and uniqueness of the Streptomycetes.

The Core Genome of Streptomycetes

We further investigated the core genome to understand the conserved basic biology of Streptomycetes. In general, *Streptomyces* species contain a linear chromosome, which has a "core region" that houses the relatively conserved housekeeping genes and two "arms" that contain more divergent and horizontally transferred genes [7]. The terminal regions of the chromosomes are highly unstable, and unequal crossing-over between the two arms of the chromosome or between one arm of the chromosome and a linear plasmid also occurs frequently, giving rise to gross rearrangements of the chromosome [7]. The dynamic nature of the arms is consistent with their high genetic diversity. Therefore, a large part of the terminal region was deleted when the genome-minimized host for heterologous expression was constructed, due to the infrequent occurrence of essential genes at the region [18]. We confirmed that there was a high frequency of core genes at the central region of the chromosome in most species, consistent with prior knowledge (Fig. 3).

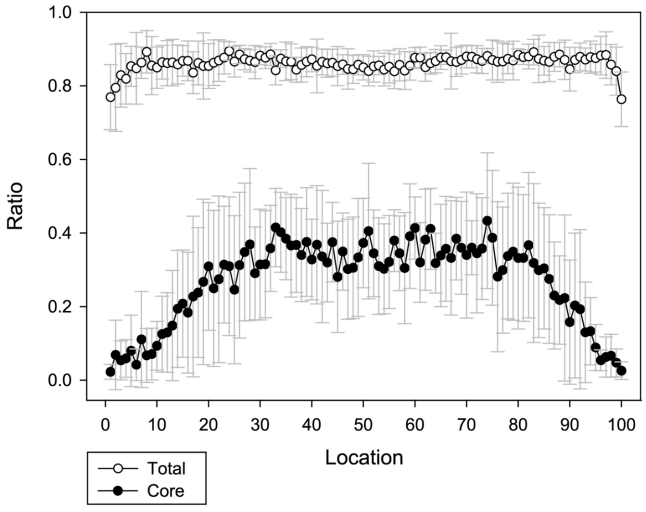

Fig. 3. Proportion of the core genome according to the location in linear chromosomes.

All genomes were normalized to the same size and divided into 100 sections. The plot represents the average ratio of the length of the total and core genes to each section. Error bars indicate the standard deviation of the ratio in each section.

Next, we further examined several groups of core genes in *S. coelicolor* as a reference strain. First, among the transcription-related genes that occupy 12% of the ORFs in the genome of *S. coelicolor*, we examined genes for sigma factors that bring diversity in the gene expression pattern by altering the specificity of RNA polymerase. The genome of *S. coelicolor* A3(2) is known to encode more than 60 different sigma factors [1, 11]. The COG clusters for sigma factors in the *Streptomyces* core genome were assigned to 15 clusters. We found that 25 out of 65 sigma factors encoded in the *S. coelicolor* genome were included in the 15 core ortholog clusters (Table 1). Genes for the major housekeeping sigma factor HrdB and its paralogs HrdA, HrdC, and HrdD were clustered in a single group (cluster ID 24). This cluster comprised 3–4 genes in each analyzed species, indicating that the multiplicity of these HrdB-like sigma factors is conserved in *Streptomyces*. Ortholog cluster 4 contained the largest number of sigma factor genes (*sigB*, *sigI*, *sigN*, *sigF*, *sigH*, and *sigL*), many of which were reported to function in differentiation and response to osmotic and oxidative stresses [30]. Except for WhiG (cluster ID 1910), all the other 14 sigma factors are classified as group 4 or ECF-family sigma factors [12]. The conserved ECF sigma factors in all *Streptomyces* spp. include SigU, SigE, SigR, SigR1, SigR, BldN, and SigQ. Among these, some are known to be involved in differentiation and secondary metabolism (SigU, BldN, SigR, and SigT) [8, 10, 20, 27], cell wall function (SigE) [25], and oxidative stress response (SigR and SigR1) [16, 17, 26]. Investigation of other conserved sigma factors is needed to unravel the conserved core functions governed by conserved alternate sigma factors.

In addition, conservation of genes involved in the biosynthesis of mycothiol, the major principal thiol compound found in many actinomycetes, was investigated [23]. This maintains a high level of reducing environment within the cells and protects against disulfide stress. Four putative genes in this pathway, *mshA* (cluster ID 705), *mshB* (cluster ID 1092), *mshC* (cluster ID 953), and *mshD* (cluster

Table 1. Conserved genes in 17 *Streptomyces* species.

Cluster ID	SCO No.	Gene name	Function
Sigma factor			
4	SCO0600	*sigB*	RNA polymerase sigma factor
4	SCO3068	*sigI*	RNA polymerase sigma factor
4	SCO4034	*sigN*	RNA polymerase sigma factor
4	SCO4035	*sigF*	RNA polymerase sigma factor
4	SCO5243	*sigH*	RNA polymerase sigma factor
4	SCO7278		RNA polymerase sigma factor
24	SCO0895	*hrdC*	RNA polymerase sigma factor
24	SCO2465	*hrdA*	RNA polymerase principal sigma factor
24	SCO3202	*hrdD*	RNA polymerase principal sigma factor
24	SCO5820	*hrdB*	RNA polymerase principal sigma factor
107	SCO0942		RNA polymerase sigma factor
107	SCO2954	*sigU*	RNA polymerase sigma factor
316	SCO4864		ECF sigma factor
316	SCO4866		ECF sigma factor
644	SCO4005		RNA polymerase sigma factor
645	SCO3356	*sigE*	ECF sigma factor
729	SCO5216	*sigR*	RNA polymerase sigma factor
882	SCO5147		RNA polymerase sigma factor
956	SCO3613		RNA polymerase sigma factor
1303	SCO4409		RNA polymerase sigma factor
1380	SCO3892	*sigT*	RNA polymerase sigma factor
1743	SCO3323	*bldN*	RNA polymerase sigma factor
1910	SCO5621	*whiG*	RNA polymerase sigma factor
2236	SCO4908	*sigQ*	RNA polymerase sigma factor
2282	SCO4769		ECF sigma factor
Mycothiol biosynthesis			
705	SCO4204	*mshA*	Putative glycosyltransferase
953	SCO1663	*mshC*	Putative cysteinyl-tRNA synthetase
1092	SCO5126	*mshB*	Conserved hypothetical protein
1601	SCO1545	*mshD*	Putative acetyltransferase

ID 1601), were conserved in all the species that we analyzed, despite of their scattered location in the chromosome (Table 1). This proved that mycothiol acts as the common reducing agent in *Streptomyces* species.

We further examined the conserved core genes that are annotated to be involved in secondary metabolism. Among genes involved in secondary metabolism, more than 95% resided mostly in the accessory (dispensable and unique) genomes of *Streptomyces* (Fig. 2). This reflects the diversity of secondary metabolism of *Streptomyces* spp. Only 5% of genes for secondary metabolism is in the conserved core genome. Table 2 lists 27 genes for secondary metabolism that are conserved among 17 *Streptomyces* spp. They belong to seven COG clusters for secondary metabolism out of 30 clusters predicted [22]. Most or all of the genes in the 5-hydroxyectoin (4/4), siderophore (2/3), geosmin (1/1), and hopene (11/13) clusters were conserved throughout *Streptomyces* spp. 5-Hydroxyectoin is known to have an important role as a compatible solute in response to salt and heat stresses in the *Streptomyces* genus [4]. Geosmin, which is responsible for the odor of soil, is also likely to be produced in all of the strains examined in this work [5]. Hopene, a pentacyclic triterpene, can provide stability to bacterial membranes at high temperatures and under conditions of extreme acidity [28]. This study reveals that among secondary metabolites, only a handful of compounds such as hydroxyectoine, geosmin, and hopanoids are universally conserved among *Streptomyces*. More intensive investigation of the functions of these metabolites, either characterized or uncharacterized, is in need to understand their roles in the biology of Streptomycetes.

The amount of bacterial genomic information has been rapidly increasing with the development of high-throughput DNA sequencing technologies. In particular, the acquisition and understanding of the genome sequences of *Streptomyces* are important for drug discovery, because these organisms are an abundant source of secondary metabolites [29]. In this study, we revealed the conservation of 2,018 and 32,574 gene families (COG clusters) within the core and accessary genomes, respectively, of 17 completely sequenced *Streptomyces* species using pan-genome analysis. Functional classification of ortholog clusters showed the distribution of ratios of core and accessory genomes. Furthermore, we investigated the functions of the conserved gene groups, which included the sigma factors, mycothiol biosynthesis pathway, and secondary metabolic pathways. This analysis showed that *Streptomyces* species encode many common genes involved in stress response and morphological differentiation. Compared with previous reports [13, 15, 32], we could reduce the number of core genes using more completed genomes. Despite of the fewer number of core genes, we could find that many genes and secondary metabolite clusters that respond to stress and external stimulus were still conserved significantly. Therefore, it is concluded that adaptation or survival in various environments is one of the distinguishing characters of *Streptomyces* genus.

Elucidation of the core genome will provide insights into target selection for genome minimization during the

Table 2. Conserved genes involved in secondary metabolism in 17 *Streptomyces* species.

Secondary metabolite	SCO No.	Gene name	Function
Coelichelin (1/11)	SCO0489	-	Hypothetical protein
5-Hydroxyectoine (4/4)	SCO1864	-	Acetyltransferase
	SCO1865	-	Diaminobutyrate−2-oxoglutarate aminotransferase
	SCO1866	*ectC*	L-Ectoine synthase
	SCO1867	*ectD*	Hydroxylase
CDA (7/40)	SCO3210	-	2-Dehydro-3-deoxyheptonate aldolase
	SCO3211	-	Indoleglycerol phosphate synthase
	SCO3212	-	Anthranilate phosphoribotransferase
	SCO3213	-	Anthranilate synthase component II
	SCO3218	-	Hypothetical protein
	SCO3221	-	Prephenate dehydrogenase
	SCO3224	-	ABC transporter ATP-binding protein
Siderophore (2/3)	SCO5800	-	Hypothetical protein
	SCO5801	-	Hypothetical protein
Geosmin (1/1)	SCO6073	-	Cyclase
Coelimycin P1 (1/16)	SCO6284	-	Decarboxylase
Hopene (11/13)	SCO6759	-	Phytoene synthase
	SCO6760	-	Phytoene synthase
	SCO6762	-	Phytoene dehydrogenase
	SCO6763	-	Polyprenyl synthatase
	SCO6764	-	Squalene-hopene cyclase
	SCO6765	-	Lipoprotein
	SCO6766	-	hypothetical protein
	SCO6767	*ispG*	4-Hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
	SCO6768	-	1-Deoxy-D-xylulose-5-phosphate synthase
	SCO6769	-	Aminotransferase
	SCO6770	-	DNA-binding protein

construction of industrial strains or for metabolic engineering. Moreover, this analysis offers a basis for understanding the processes through which information from one strain is transferred to another strain. Integration of genomic information with other -omics studies, such as transcriptomics, proteomics, and metabolomics, will provide an opportunity for understanding more about the functional evolution of *Streptomyces* species.

Acknowledgments

This work was supported by the Intelligent Synthetic Biology Center of the Global Frontier Project (2011-0031957) and the Basic Core Technology Development Program for the Oceans and the Polar Regions (2011-0021053) through the National Research Foundation of Korea (NRF), which is funded by the Ministry of Science, ICT, and Future Planning.

References

1. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, *et al.* 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417:** 141-147.

2. Bibb MJ. 2005. Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.* **8:** 208-215.
3. Binnewies TT, Motro Y, Hallin PF, Lund O, Dunn D, La T, *et al.* 2006. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct. Integr. Genomics* **6:** 165-185.
4. Bursy J, Kuhlmann AU, Pittelkow M, Hartmann H, Jebbar M, Pierik AJ, Bremer E. 2008. Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3(2) in response to salt and heat stresses. *Appl. Environ. Microbiol.* **74:** 7286-7296.
5. Cane DE, Watt RM. 2003. Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. *Proc. Natl. Acad. Sci. USA* **100:** 1547-1551.
6. Chandra G, Chater KF. 2014. Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* **38:** 345-379.
7. Dyson P. 2011. *Streptomyces: Molecular Biology and Biotechnology*. Horizon Scientific Press Norfolk, UK.
8. Feng WH, Mao XM, Liu ZH, Li YQ. 2011. The ECF sigma factor SigT regulates actinorhodin production in response to nitrogen stress in *Streptomyces coelicolor*. *Appl. Microbiol. Biotechnol.* **92:** 1009-1021.
9. Flärdh K, Buttner MJ. 2009. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7:** 36-49.
10. Gehring AM, Yoo NJ, Losick R. 2001. RNA polymerase sigma factor that blocks morphological differentiation by *Streptomyces coelicolor*. *J. Bacteriol.* **183:** 5991-5996.
11. Hahn MY, Bae JB, Park JH, Roe JH. 2003. Isolation and characterization of *Streptomyces coelicolor* RNA polymerase, its sigma, and antisigma factors. *Methods Enzymol* **370:** 73-82.
12. Helmann JD. 2002. The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.* **46:** 47-110.
13. Hsiao NH, Kirby R. 2007. Comparative genomics of *Streptomyces avermitilis*, *Streptomyces cattleya*, *Streptomyces maritimus* and *Kitasatospora aureofaciens* using a *Streptomyces coelicolor* microarray system. *Antonie Van Leeuwenhoek* **93:** 1-25.
14. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, *et al.* 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21:** 526-531.
15. Jayapal KP, Lian W, Glod F, Sherman DH, Hu W-S. 2007. Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. *BMC Genomics* **8:** 229.
16. Kallifidas D, Thomas D, Doughty P, Paget MS. 2010. The sigmaR regulon of *Streptomyces coelicolor* A32 reveals a key role in protein quality control during disulphide stress. *Microbiology* **156:** 1661-1672.
17. Kim M-S, Dufour YS, Yoo JS, Cho Y-B, Park J-H, Nam G-B, *et al.* 2012. Conservation of thiol-oxidative stress responses regulated by SigR orthologues in actinomycetes. *Mol. Microbiol.* **85:** 326-344.
18. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. 2010. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. USA* **107:** 2646-2651.
19. Liu G, Chater KF, Chandra G, Niu G, Tan H. 2013. Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **77:** 112-143.
20. Mao XM, Zhou Z, Cheng LY, Hou XP, Guan WJ, Li YQ. 2009. Involvement of SigT and RstA in the differentiation of *Streptomyces coelicolor*. *FEBS Lett.* **583:** 3145-3150.
21. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15:** 589-594.
22. Nett M, Ikeda H, Moore BS. 2009. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26:** 1362-1384.
23. Newton GL, Buchmeier N, Fahey RC. 2008. Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **72:** 471-494.
24. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, *et al.* 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **190:** 4050-4060.
25. Paget MS, Chamberlin L, Atrih A, Foster SJ, Buttner MJ. 1999. Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181:** 204-211.
26. Paget MS, Kang JG, Roe JH, Buttner MJ. 1998. sigmaR, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). *EMBO J.* **17:** 5776-5782.
27. Paget MS, Molle V, Cohen G, Aharonowitz Y, Buttner MJ. 2001. Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol. Microbiol.* **42:** 1007-1020.
28. Poralla K, Muth G, Hartner T. 2000. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* **189:** 93-95.
29. Rebets Y, Brotz E, Tokovenko B, Luzhetskyy A. 2014. Actinomycetes biosynthetic potential: how to bridge *in silico* and *in vivo*? *J. Ind. Microbiol. Biotechnol.* **41:** 387-402.
30. Viollier PH, Kelemen GH, Dale GE, Nguyen KT, Buttner MJ, Thompson CJ. 2003. Specialized osmotic stress response systems involve multiple SigB-like sigma factors in *Streptomyces coelicolor*. *Mol. Microbiol.* **47:** 699-714.
31. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: pan-genomes analysis pipeline. *Bioinformatics.* **28:** 416-418.
32. Zhou Z, Gu J, Li Y-Q, Wang Y. 2012. Genome plasticity and systems evolution in *Streptomyces*. *BMC Bioinformatics.* **13 (Suppl. 10):** S8.