Bibliography

Ahn, J. H., Lee, S. A., Kim, J. M., Kim, M. S., Song, J. and Weon, H. Y.,(2016). Dynamics of bacterial communities in rice field soils as affected by different long-term fertilization practices. *J. Microbiol.,* 54(11), pp. 724-731.

Ahrens, R. and Moll, G. (1970). A new budding bacterium from the Baltic Sea. *Arch. Microbiol*., 70(3), 243.

Ahsan, T., Chen, J., Zhao, X., Irfan, M. and Wu, Y., (2017). Extraction and identification of bioactive compounds (eicosane and dibutyl phthalate) produced by Streptomyces strain KX852460 for the biological control of Rhizoctonia solani AG-3 strain KX852461 to control target spot disease in tobacco leaf. *AMB Express*, 7(1), pp. 1-9.

Ajilogba, C. F. and Babalola, O. O., (2019). GC–MS analysis of volatile organic compounds from Bambara groundnut rhizobacteria and their antibacterial properties. *World J. Microbiol. Biotechnol.* 35(6), pp. 1-19.

Akashi, H. and Gojobori, T., (2002). Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Pros Nat. Acad. Sc.,* 99(6), pp. 3695-3700.

Albarano, L., Esposito, R., Ruocco, N. and Costantini, M.,(2020). Genome mining as new challenge in natural products discovery. *Marine drugs*, 18(4), p. 199.

Aldesuquy, H. S., Mansour, F. A. and Abo-Hamed, S. A., (1998). Effect of the culture filtrates of Streptomyces on growth and productivity of wheat plants. *Folia Microbiol.,* 43(5), pp. 465-470.

Almendro-Candel, M. B., Lucas, I. G., Navarro -Pedreño, J. and Zorpas, A. A., (2018). Physical properties of soils affected by the use of agricultural waste. *Agricultural waste and residues*, pp. 9-27.

Alori, E. T., Adekiya, A. O. and Adegbite, K. A., (2020). Impact of Agricultural Practices on Soil Health. *Soil Health*, pp. 89-98.

Amin, D. H., Abdallah, N. A., Abolmaaty, A.Tolba, S. and Wellington, E. M., (2020). Microbiological and molecular insights on rare Actinobacteria harboring bioactive prospective. *Bullet. Nat. Res. Centre,* 44(1), pp. 1-12.

Anandan, R., Dharumadurai, D. and Manogaran, G. P. (2016). An introduction to actinobacteria. In Actinobacteria-Basics and Biotechnological Applications. Intechopen.

Aouiche, A., Sabaou, N., Meklat, A., Zitouni, A., Bijani, C., Mathieu, F. and Lebrihi, A. (2012) Saccharothrix sp. PAL54, a new chloramphenicol-producing strain isolated from a Saharan soil. *World J. Microbiol. Biotechnol.,* 28(3), 943-951.

Arasu, M. V., Duraipandiyan, V. and Ignacimuthu, S., 2013. Antibacterial and antifungal activities of polyketide metabolite from marine Streptomyces sp. AP-123 and its cytotoxic effect. *Chemosphere*, *90*(2), pp. 479-487.

Arroyo-Rodríguez, V., Rös, M., Escobar, F., Melo, F. P., Santos, B. A., Tabarelli, M. and

Chazdon, R., 2013. Plant β-diversity in fragmented rain forests: testing floristic homogenization and differentiation hypotheses. *J. Ecology*, 101(6), pp. 1449-1458.

Atta, H. M., (2015). Biochemical studies on antibiotic production from Streptomyces sp. : Taxonomy, fermentation, isolation and biological properties. *J. Saudi Chem. Society*, 19(1), pp. 12-22.

Axenov-Gibanov DV, Voytsekhovskaya IV, Tokovenko BT, Protasov ES, Gamaiunov SV, Rebets YV, et al. (2016). Actinobacteria Isolated from an Underground Lake and Moonmilk Speleothem from the Biggest Conglomeratic Karstic Cave in Siberia as Sources of Novel Biologically Active Compounds. *PLoS ONE* 11(2): e0149216.

Badji, B., Zitouni, A., Mathieu, F., Lebrihi, A. and Sabaou, N., 2006. Antimicrobial compounds produced by Actinomadura sp. AC104 isolated from an Algerian Saharan soil. *Canadian J. Microbiol.,* 52(4), pp. 373-382.

Badji, B. ; Mostefaoui, A. ; Sabaou, N. ; Lebrihi, A. ; Mathieu, F. ; Seguin, E. ; Tillequin, F. (2007). Isolation and partial characterization of antimicrobial compounds from a new strain Nonomuraea sp. NM94. *J. Ind. Microbiol. Biotechnol.* 2007, 34, 403–412.

Bae, E. A., Han, M. J., Song, M. J. and Kim, D. H., (2002). Purification of rotavirus infection-inhibitory protein from Bifidobacterium breve K-110. *J. Microbiol. Biotechnol,* 12(4), pp. 553-556.

Bai, Y., Yang, D., Wang, J., Xu, S., Wang, X. and An, L. (2006). Phylogenetic diversity of culturable bacteria from alpine permafrost in the Tianshan Mountains, northwestern China. *Res. Microbiol*, 157(8), pp. 741-751.

Balachandran, C., Duraipandiyan, V., Emi, N. and Ignacimuthu, S. (2015). Antimicrobial and cytotoxic properties of Streptomyces sp. (ERINLG-51) isolated from Southern Western Ghats. *South Ind. J. Biol. Sci,* 1 (714), p. 5.

Bapatla, K. G., Kumar, K., Kumar, L., Singh, B. and Singh, N. P., 2021. Anti-larval activity of actinobacterial extract for Helicoverpa armigera and Spodoptera litura. *Int. J. Tropical Insect Sci.,* pp. 1-11.

Baranasic, D., Gacesa, R., Starcevic, A., Zucko, J., Blažič, M., Horvat, M., Gjuračić, K., Fujs, Š., Hranueli, D., Kosec, G. and Cullum, J. (2013). Draft genome sequence of Streptomyces rapamycinicus strain NRRL 5491, the producer of the immunosuppressant rapamycin. *Genome Announc*., 1(4): e00581-13. doi: 10. 1128/*GenomeA*. 00581-13

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H. P., Clément, C., Ouhdouch, Y. and van Wezel, G. P., 2016. Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol. Mol. Biol. Reviews*, *80*(1), pp. 1-43.

Baruah, T. C. and Barthakur, H. P., (1997). A textook of soil chemical analysis. Vikash, New Delhi.

Bauer, A. W.,(1966.) Antibiotic susceptibility testing by a standardized single disc method. *Am J Clin Pathol,* 45, pp. 149-158.

Behura, S. K. and Severson, D. W.,(2013). Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biological Reviews*, 88(1), pp. 49-61.

Bennett, S., 2004. Solexa ltd. Pharmacogenomics, 5(4), pp. 433-438.

Bentley, S. D., Maiwald, M., Murphy, L. D., Pallen, M. J., Yeats, C. A., Dover, L. G., Norbertczak, H. T., Besra, G. S., Quail, M. A., Harris, D. E. and von Herbay, A. (2003). Sequencing and analysis of the genome of the Whipple's disease bacterium Tropheryma whipplei. *The Lancet*, 361(9358), pp. 637-644.

Berd, D., (1973). Laboratory identification of clinically important aerobic actinomycetes. *Applied Microbiol*., 25(4), pp. 665-681.

Berdy J (2005). Bioactive microbial metabolites. *J. Antibiot* 58:1

Berg, G. and Smalla, K., 2009. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol. Ecol*., 68(1), pp. 1-13.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M. C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H. and

Kazou, M., 2020. Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, 8(1), pp. 1-22.

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., Knight, R. and Fierer, N., (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol. Biochem*., 43(7), pp. 1450-1455.

*Bernard, G.,and. DromardA (2011); Book of etymology and medical terminology: Lexicon etymology. Livretd'étymologieet de terminologie médicale: Lexiqued' étymologie, pp. 1-4.

Bhatti, A. A., Haq, S. and Bhat, R. A., (2017.) Actinomycetes benefaction role in soil and plant health. *Microbial pathogenesis*, 111, pp. 458-467.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H. and Weber, T., (2019.) antiSMASH 5. 0: updates to the secondary metabolite genome mining pipeline. *Nucleic acids Res.,* 47(W1), pp. W81-W87.

Bodilis, J. and Barray, S., (2006). Molecular evolution of the major outer-membrane protein gene (oprF) of Pseudomonas. *Microbiol*., 152(4), pp. 1075-1088.

Bolger, A. M., Lohse, M. and Usadel, B., (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30 (15), pp. 2114-2120.

Borah, A. and Thakur, D., (2020). Phylogenetic and functional characterization of culturable endophytic actinobacteria associated with Camellia spp. for growth promotion in commercial tea cultivars. *Front. Microbiol*., 11, p. 318.

Borborah S. & J. K Gogoi (2007) "Development issues in the North-East in the Turn of the country", published Papyrus Books, Guwahati, Assam

Borovska, P. and Ivanova, D., 2014. Code optimization and scaling of the astrophysics software Gadget on Intel Xeon Phi. Partnership for Advanced Computing in Europe (PRACE), *136*.

Bottone, E. J., (2010). Bacillus cereus, a volatile human pathogen. *Clinical Microbiol*. Rev., 23(2), pp. 382-398.

*Bream, A. S., Ghazal, S. A., El-Aziz, Z. K. A. and Ibrahim, S. Y., 2001. Insecticidal activity of selected actinomycetes strains against the Egyptian cotton leaf worm Spodoptera littoralis (Lepidoptera: Noctuidae). Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen Universiteit Gent, *66*(2a), pp. 503-544.

Broeckling, C. D., Broz, A. K., Bergelson, J., Manter, D. K. and Vivanco, J. M., (2008). Root exudates regulate soil fungal community composition and diversity. *Applied and Environ. Microbiol.,* 74(3), pp. 738-744.

Bull A. T. (2011) Actinobacteria of the Extremobiosphere. In: Horikoshi K. (eds) Extremophiles Handbook. Springer, Tokyo. https://doi. org/10. 1007/978-4-431-53898-1_58

*Bull AT, Stach JE, Ward AC, Goodfellow M. (2005) Marine Actinobacteria: perspectives, challenges, future directions. A. Van. Leeuw. ; 87(1):65-79.

Bushnell, B., (2014). BBMap: a fast, accurate, splice-aware aligner (No. LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Caesar-TonThat, T. C. and Cochran, V. L., (2000). Soil aggregate stabilization by a saprophytic lignin-decomposing basidiomycete fungus I. Microbiological aspects. Biol. and fertility of soils, 32(5), pp. 374-380. Cappuccino, J. G. and Sherman, N., 2008. Microbiol. : a laboratory manual (Vol. 9).

Cappuccino, J. G. and Sherman, N., (1992). Biochemical activities of microorganisms. Microbiol., A Laboratory Manual. The Benjamin/Cummings Publishing Co. California, USA, pp. 188-247.

Cary, S. C., McDonald, I. R., Barrett, J. E. and Cowan, D. A. (2010). On the rocks: the Microbiol. of Antarctic Dry Valley soils. *Nat. Rev. Microbiol.*, 8(2), pp. 129-138.

Chapman TM, Perry CM (2004). Everolimus. Drugs 64:861-872

Chapman, H. D. and Pratt, P. F., 1961. Methods of analysis for soils. *Plants and Waters*, pp. 169 -176.

Chater, K. F. and Chandra, G. (2006). The evolution of development in Streptomyces analysed by genome comparisons. *FEMS Microbiol Rev*, 30(5), pp. 651-672.

Chen, Z. M. and Chen, X. F., (1989). An analysis of world tea pest fauna. *J. Tea Sci*, 9 (1), pp. 13-22.

Chenoll, E., Rivero, M., Codoñer, F. M., Martinez-Blanch, J. F., Ramón, D., Genovés, S. and Moreno Muñoz, J. A., (2015). Complete genome sequence of Bifidobacterium longum subsp. infantis strain CECT 7210, a probiotic strain active against rotavirus infections. *GenomeA*, 3(2), pp. e00105-15.

Chevrette, M. G., Carlos-Shanley, C., Louie, K. B., Bowen, B. P., Northen, T. R. and Currie, C. R., 2019. Taxonomic and metabolic incongruence in the ancient genus Streptomyces. . *Front. Microbiol.,* p. 2170.

Cho S-H, Han J-H, Ko H-Y, Kim SB (2008) Streptacidiphilus anmyonensis sp. nov., Streptacidiphilus rugosus sp. nov. and Streptacidiphilus melanogenes sp. nov., acidophilic acti-nobacteria isolated from Pinus soils. *Int J Syst Evol Microbiol* 58:1566–1570

Chowdhury, A., Sarkar, J., Chakraborti, T., Pramanik, P. K. and Chakraborti, S., 2016. Protective role of epigallocatechin-3-gallate in health and disease: a perspective. Biomedicine & Pharmacotherapy, 78, pp. 50 -59.

Comeron, J. M. and Aguadé, M., (1998). An evaluation of measures of synonymous codon usage bias. *J. Mol. Evolution*, 47(3), pp. 268-274.

Corbaz, R., Gregory, P. H., and lacey, M. E. (1963). Thermophilic and mesophilic actinomycetes in mouldy hay. *J. Gen. Microbiol.* 32, 449–455. doi: 10. 1099/00221287-32-3-449

Cortez D., Delaye L., Lazcano A., Becerra A. Horizontal Gene Transfer. Totowa, New Jersey, USA: Humana Press; (2009). Composition-based methods to identify horizontal gene transfer; pp. 215–25.

Costa, O. Y., Raaijmakers, J. M. and Kuramae, E. E., (2018). Microbial extracellular polymeric substances: ecological function and impact on soil aggregation. *Front. Microbiol*., 9, p. 1636.

Costa, R., Salles, J. F., Berg, G. and Smalla, K., (2006). Cultivation-independent analysis of Pseudomonas species in soil and in the rhizosphere of field-grown Verticillium dahliae host plants. *Environmental Microbiol.,* 8(12), pp. 2136-2149.

Costa, S. S., Guimarães, L. C., Silva, A., Soares, S. C. and Baraúna, R. A., (2020). First steps in the analysis of prokaryotic pan-genomes. Bioinformatics and Biol. Insights, 14, p. 1177932220938064.

Cragg GM, Newman DJ (2009). Nature: a vital source of leads for anticancer drug development. *Phytochem. Rev.* 8:313-331

Cross T. (1981). Aquatic actinomycetes: a critical survey of the occurrence, growth and role of actinomycetes in aquatic habitats. *J Appl Bacterriol.* 1981; 50(3):397-423.

Culver, D. C. and Sket, B. (2000). Hotspots of subterranean biodiversity in caves and wells. *J. Cav. Kar. Stud*., 62(1), pp. 11-17.

Cundell DR and Piechoski MP (2016) Potentially novel Actinobacteria derived antibiotics from unique microenvironments. In: Antimicrobials Synthetic and Natural Compounds. (Dhanasekaran et al., Eds), 2016, pp: 83-98, CRC Press, New York.

Curtis, T. P. and Sloan, W. T., (2005). Exploring microbial diversity--a vast below. *Science.,* 309(5739), pp. 1331-1333.

da C Jesus, E., Marsh, T. L., Tiedje, J. M. and de S Moreira, F. M., (2009). Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J.*, 3(9), pp. 1004-1011.

Daniel, R., (2005). The metagenomics of soil. *Nature Rev. Microbiol*., 3(6), pp. 470-478.

de Caire, G. Z., De Cano, M. S., De Mule, M. Z., Palma, R. M. and Colombo, K., (1997). Exopolysaccharide of Nostoc muscorum (Cyanobacteria) in the aggregation of soil particles. *J. Applied Phycology*, 9(3), pp. 249-253.

de Jesus Sousa, J. A. and Olivares, F. L., (2016). Plant growth promotion by streptomycetes: ecophysiology, mechanisms

and applications. Chemical and Biological *Technologies in Agriculture*, 3(1), pp. 1-12

Dechner, A., Flesher, K. M., Lindell, C., Vega de Oliveira, T. and Maurer, B. A., (2018.) Determining carnivore habitat use in a rubber/forest landscape in Brazil using multispecies occupancy models. *PloS one,* 13(4), p. e0195311.

Deltedesco, E., Keiblinger, K. M., Piepho, H. P., Antonielli, L., Pötsch, E. M., Zechmeister -Boltenstern, S. and Gorfer, M., (2020). Soil microbial community structure and function mainly respond to indirect effects in a multifactorial climate manipulation experiment. *Soil Biol. Biochem.*, 142, p. 107704.

Demain, A. L. and Sanchez, S. (2009). Microbial drug discovery: 80 years of progress. *J. Antibiot*., 62(1), pp. 5-16.

Dhanasekaran D, Panneerselvam A, Thajuddin N, Chandralekha S. (2014). Isolation, characterization of antibacterial methyl substituted β-lactam compound from Streptomyces noursei DPTD21 in saltpan soil, India. *JBAPN*.; 4(2):71-88.

Dhanasekaran D, Rajakumar G, Sivamani P, Selvamani S, Panneerselvam A, Thajuddin N. (2005). Screening of salt pans actinomycetes for antibacterial agents. *Int. J Microbiol.;* 1(2):1-8.

Dhanasekaran D, Sakthi V, Thajuddin N, Panneerselvam A. (2010). Preliminary evaluation of Anopheles mosquito larvicidal efficacy of mangrove Actinobacteria. *Int J Appl Biol Pharm Technol*.; 1(2):374-381.

Ding, D., Chen, G., Wang, B., Wang, Q., Liu, D., Peng, M. and Shi, P. (2013). Culturable actinomycetes from desert ecosystem in northeast of Qinghai-Tibet Plateau. *Ann. Microbiol*, 63(1), pp. 259-266.

*Diniz, A. R., Silva, C. F. D., Pereira, M. G., Balieiro, F. C., Silva, E. V. D. and Santos, F. M. D., (2020). Microbial Biomass and Enzyme Activity of Soil Under Clonal Rubber Tree Plantations. *Floresta e Ambiente*, 27.

Direito, S. O., Ehrenfreund, P., Marees, A., Staats, M., Foing, B. and Röling, W. F. (2011). A wide variety of putative extremophiles and large beta-diversity at the Mars Desert Research Station (Utah). *Internat. J. Astrobiol*., 10(3), p. 191.

Dong, Y., Lin, H., Zhao, Y. and Menzembere, E. R. G. Y., (2021). Remediation of vanadium-contaminated soils by the combination of natural clay mineral and humic acid. J. Cleaner Production, 279, p. 123874.

Doroghazi, J. R, Metcalf W. W (2013). Comparative genomics of actinobacteria with a focus on natural product biosynthetic genes. *BMC Genom.* 14:611

Dos Reis, M., Wernisch, L. and Savva, R., (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. *Nucleic acids Res*., 31(23), pp. 6976-6985.

Effendi, Y., Aini, N., Pambudi, A. and Sasaerila, H. Y., (2020), March. Metagenomics analysis of soil microbial communities in plant agroforestry system rubber tree (Hevea brasiliensis)–Ganyong (Canna sp.). In IOP Conference Series: Earth and Environmental Sci. (Vol. 468, No. 1, p. 012045). IOP Publishing.

Eichorst, S. A., Trojan, D., Roux, S., Herbold, C., Rattei, T. and Woebken, D., (2018). Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. *Environmental Microbiol*., 20 (3), pp. 1041-1063.

Elbendary, A. A., Hessain, A. M., El-Hariri, M. D., Seida, A. A., Moussa, I. M., Mubarak, A. S., Kabli, S. A., Hemeg, H. A. and El Jakee, J. K., (2018). Isolation of antimicrobial producing actinobacteria from soil samples. *Saudi J. biological Sci*., 25(1), pp. 44-46.

El-Khawagh, M. A., Hamadah, K. S. and El-Sheikh, T. M., 2011. The insecticidal activity of actinomycete metabolites, against the mosquito Culex pipiens. *Egypt Acad J Biol Sci*, *4*(1), pp. 103-113.

El-Naggar, N. E. A. and El-Ewasy, S. M., (2017). Bioproduction, characterization, anticancer and antioxidant activities of extracellular melanin pigment produced by newly isolated microbial cell factories Streptomyces glaucescens NEAE-H. *Scientific Rep.,* 7(1), pp. 1-19.

Eppard, M., Krumbein, W. E., Koch, C., Rhiel, E., Staley, J. T. and Stackebrandt, E. (1996). Morphological, physiological, and molecular characterization of actinomycetes isolated from dry soil, rocks, and monument surfaces. *Arch. Microbiol*., 166(1), pp. 12-22.

Erikstad, H. A. and Birkeland, N. K., (2015). Draft genome sequence of "Candidatus Methylacidiphilum kamchatkense" strain Kam1, a thermoacidophilic methanotrophic verrucomicrobium. *GenomeA,* 3(2), pp. e00065-15.

Ermolaev, M. D., (2001). Synonymous codon usage in bacteria. *Current issues in Mol. Biol*., 3(4), pp. 91-97.

Essaid Ait Barka, Parul Vatsa, Lisa Sanchez, Nathalie GaveauVaillant, Cedric Jacquard, HansPeter Klenk, Christophe Clément, Yder Ouhdouch, Gilles P. van Wezel (2015). Taxonomy, Physiology, and Natural Products of Actinobacteria. Microbiol. Mol. Biol. Rev. 80 (1) 1-43; DOI: 10. 1128/ *MMBR*. 00019-15

Fanani, A. N., Rahardja, B. S. and Prayogo, P., (2018). Efek Padat Tebar Ikan Lele Dumbo (Clarias SP.) yang Berbeda terhadap Kandungan Amonia (NH3) dan Nitrit (NO2) dengan Sistem Bioflok. *J. Aquaculture Sci*., 3(2), p. 276555.

Faust, K. and Raes, J.,(2012.) Microbial interactions: from networks to models. *Nature Rev.Microbiol*., 10(8), pp. 538-550.

Finna, P., Usman, P., Wahyu, P. N. and Aulia, A., (2019,) May. Isolation and Identification of Bacteria and Actinomycetes Isolated from Wilting Banana Plants (Musa Sp.). In IOP Conference Series: Materials Sci. and Engineering (Vol. 532, No. 1, p. 012028). IOP Publishing.

Frost, F. J., Tollestrup, K., Roberts, M., Kunde, T. R., Craun, G. F. and Harter, L., (2009). Enteric illness risks before and after water treatment improvements. *J. Water Health*, 7(4), pp. 581-589.

Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. and Garrels, J. I., (1999). A sampling of the yeast proteome. *Mol. Cellular Biol*., 19(11), pp. 7357-7368.

G Rajput, S., Hense, S. and Thankappan, K. R., (2021). Healthcare utilisation: a mixed-method study among tea garden workers in Indian context. J. Health Research. ayathri, M. P. and Arjunan, R., Health Afflictions of Tea Plantation Workers in Coonoor, the Nilgiris.

Gadelhak, G. G., El-Tarabily, K. A. and Al-Kaabi, F. K.,(2005). Insect control using chitinolytic soil actinomycetes as biocontrol agents. *Int J Agri. Biol.* 7(4), pp. 627-633.

Gao, S., DeLuca, T. H. and Cleveland, C. C., (2019). Biochar additions alter phosphorus and nitrogen availability in agricultural ecosystems: a meta-analysis. *Sci. Total Environ.*, 654, pp. 463-472.

Garbeva, P. V., Van Veen, J. A. and Van Elsas, J. D., (2004). Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. *Annu. Rev. Phytopathol.,* 42, pp. 243-270.

Gardner, H. L. and Dukes, C. D. (1955) Haemophilus vaginalis vaginitis: a newly defined specific infection previously classified "nonspecific" vaginitis. *Am. J. Obstr. Gynecol*., 69(5), pp. 962-976.

Gayathri, M. P. and Arjunan, R., 2019. Health Ailments of Tea Plantation Workers in Coonoor, The Nilgiris. *Think India J.,* 22(4), pp. 6633-6638.

Ghazal, S. A., Bream, A. S., Abd el-Aziz, Z. K. and Ibrahim, S. Y., 2001. Preliminary studies on insecticidal activities of actinomycete strains propagated on solid and broth media using Musca domestica (Diptera: Muscidae). *Meded Rijksuniv Gent Fak Landbouwkd Toegep Biol Wet*, *66*(2b), pp. 559-70.

Nawani, N., Aigle, B., Mandal, A., Bodas, M., Ghorbel, S. and Prakash, D., 2013. Actinomycetes: Role in biotechnology and medicine. *BioMed Res. Int*., 2013.

Ghosh, S., (2016). Selling Nature: Narratives of Coercion. Resistance and Ecology. Business Interests and Environmental Crisis. Sage. Delhi, pp. 170-206.

Ghosh, U., Haq, M. B. and Chakrabarty, S. (2011). Application of systematic technologies for the extraction of novel phytoconstituents from pharmacologically important plants. *Int. J. Chem. Analyt. Sci*, 2, pp. 1153-1158.

Girão, M., Ribeiro, I., Ribeiro, T., Azevedo, I. C., Pereira, F., Urbatzka, R., Leão, P. N. and Carvalho, M. F., (2019). Actinobacteria isolated from Laminaria ochroleuca: a source of new bioactive compounds. *Front. Microbiol.,* 10, p. 683.

Gökçeoğlu, M., 1988. Nitrogen mineralization in volcanic soil under grassland, shrub and forest vegetation in the Aegean region of Turkey. *Oecologia*, *77*(2), pp. 242-249.

Goodfellow M, Williams ST (1983) Ecology of actinobacteria. *Annual Rev. Microbiol.* 37:189-216

Goodfellow, M., Haynes, J. A. (1984). Actinobacteria in marine sediments. Biological, biochemical and biomedical aspects of Actinobacteria: 453-472

Gopinath, S. C., Anbu, P., Arshad, M. M., Lakshmipriya, T., Voon, C. H., Hashim, U. and Chinni, S. V.,(2017). Biotechnological processes in microbial amylase production. *BioMed Res. Int*., 2017.

Gordon, R. E. and Smith, M. M., (1955). Proposed group of characters for the separation of Streptomyces and Nocardia. J. *Bact.,* 69(2), pp. 147-150.

Grady, F. (2005). Collecting in caves. Vertebrate Paleontological Techniques. Cambridge University Press, 2005.

Grossart H. P, Schlingloff A, Bernhard M, Simon M, Brinkhoff T. (2004). Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea. *FEMS Microbial. Ecol*. ; 47(3):387-396.

Gtari, M., Essoussi, I., Maaoui, R., Sghaier, H., Boujmil, R., Gury, J., Pujic, P., Brusetti, L., Chouaia, B., Crotti, E. and Daffonchio, D. (2012). Contrasted resistance of stone-dwelling Geodermatophilaceae species to stresses known to give rise to reactive oxygen species. *FEMS Microbiol. Ecol*., 80 (3), pp. 566-577.

Guo, H. C., Wang, W. B., Luo, X. H. and Wu, X. P., (2015). Characteristics of rhizosphere and bulk soil microbial communities in rubber plantations in Hainan Island, China. *J. Tropical Forest Sci*., pp. 202-212.

Gurusubramanian, G., Kumar, N. S., Tamuli, A. K., Sarmah, M., Rahman, A., Bora, S. and Roy, S., (2008). Biointensive Integrated Management of tea pests for sustainable tea production in North East India. *Int. J. Tea Sci.,* 7(03 and 04), pp. 45-59.

Gygi, S. P., Rochon, Y., Franza, B. R. and Aebersold, R., (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cellular Biol.*, 19(3), pp. 1720-1730.

Hamdali, H., Virolle, M. J., von Jan, M., Spröer, C., Klenk, H. P. and Ouhdouch, Y., (2011). Streptomyces youssoufiensis sp. nov., isolated from a Moroccan phosphate mine. *Int. J. system. Evol. Microbiol.,* 61(5), pp. 1104-1108

Hao, B. Z. and Wu, J. L.,(2000). Laticifer differentiation in Hevea brasiliensis: induction by exogenous jasmonic acid and linolenic acid. *Annals Bot.*, 85(1), pp. 37-43.

Hassink, J., Bouwman, L. A., Zwart, K. B., Bloem, J. and Brussaard, L., (1993). Relationships between soil texture, physical protection of organic matter, soil biota, and C and N mineralization in grassland soils. In Soil Structure/Soil Biota Interrelationships (pp. 105-128). Elsevier.

Hayakawa, T., Otake, N., Yonehara, Y., Tanaka, T. and Sakaguchi, K. (1979). Isolation and characterization of plasmids from Streptomyces. *J. Antibiot* 32(12), pp. 1348-1350.

Haynes, R. J. and Beare, M. H., (2020). Aggregation and organic matter storage in meso-thermal, humid soils. In Structure and organic matter storage in agricultural soils (pp. 213-262). CRC Press.

Hazarika, K. and Borah, K., 2013. Small tea cultivation in the process of selfemployment: a study on the indigenous people of Assam (India).

Hein, J. W., Wolfe, G. V. and Blee, K. A., (2008). Comparison of rhizosphere bacterial communities in Arabidopsis thaliana mutants for systemic acquired resistance. *Microbial. Ecol.,* 55(2), pp. 333-343.

Helmke E, Weyland H (1984) *Rhodococcus marinonascens* sp. nov. an actinomycete from the sea. *Int. J. system. Evol. Microbiol*. 34:127-138

Henssen, A., and Schnepf, E. (1967). Zur

Kenntnisthermophiler Actinomyceten. *Arch. Mikrobiol.* 57, 214–231. doi:10. 1007/ BF00405948

*Hiltner L. and Stormer, K. (1903). Studied über die Bakterienflora des Ackerbodens, mit besonderer Berücksichtigung ihres Verhaltens nach einer Behandlung mit Schwefelkohlenstoff und nach Brache. Arb. Biol. Reichsanst Land-Forstwirtsch 3:443–545. (In German.)

Hopwood, D. A. (2006). Soil to genomics, the Streptomyces chromosome. *Ann. Rev. Genet*. 40, 1–23. doi: 10. 1146/annurev. genet. 40. 110405. 090639

Horikoshi, K. (1999). Alkaliphiles: some applications of their products for biotechnology. *Microbiol Mol Biol Rev* 63:735-750

Hucker, G. J. and Conn, H. J., 1923. Methods of Gram staining.

Hunter, E. M., Mills, H. J. and Kostka, J. E., 2006. Microbial community diversity associated with carbon and nitrogen cycling in permeable shelf sediments. *Appl. Environ. Microbiol., 72*(9), pp. 5689-5701.

Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol*. (1985);2:13–34.

Ikemura, T., (1981). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol*., 151(3), pp. 389-409.

Inagaki F, Suzuki M, Takai K, Oida H, Sakamoto T, Aoki K, Nealson K. H, Horikoshi K. (2003). Microbial communities associated with geological horizons in coastal sub seafloor sediments from the Sea of Okhotsk. *Appl Environ Microbiol.,* 69 (12):7224-7235.

Inglis, T. J. and Sagripanti, J. L., (2006). Environmental factors that affect the survival and persistence of Burkholderia pseudomallei. *Appl. Environ. Microbiol*., 72 (11), pp. 6865-6875.

Jackson, M. L., (2005). Soil chemical analysis: advanced course. UW-Madison Libraries Parallel Press.

Jacob, J., Disnar, J. R. and Bardoux, G., (2008). Carbon isotope evidence for sedimentary miliacin as a tracer of Panicum miliaceum (broomcorn millet) in the sediments of Lake le Bourget (French Alps). *Org. Geochem*. 39 (8), pp. 1077-1080.

Jacoby, R., Peukert, M., Succurro, A., Koprivova, A. and Kopriva, S., (2017). The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. *Front. plant Sci*., 8, p. 1617.

Jain, R. and Saxena, J., 2019. Impact assessment of microbial formulations in agricultural soil. In Microbial Interventions in Agriculture and Environment (pp. 471-495). *Springer*, Singapore.

Jallow, M. F., Awadh, D. G., Albaho, M. S., Devi, V. Y. and Thomas, B. M., 2017. Pesticide knowledge and safety practices among farm workers in Kuwait: Results of a survey. *Int. J. Environ. Res. public health*, 14 (4), p. 340.

Jangid, K., Williams, M. A., Franzluebbers, A. J., Sanderlin, J. S., Reeves, J. H., Jenkins, M. B., Endale, D. M., Coleman, D. C. and Whitman, W. B., (2008). Relative impacts land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biol. Biochem*., 40(11), pp. 2843-2853

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F. and Gerstein, M., (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*., 302(5644), pp. 449-453.

Jensen PR, Lauro FM (2008). An assessment of actinobacterial diversity in the marine environment. *Anton. V. Leeuwenhoek* 94:51-62

Jensen PR, Mincer TJ, Williams PG, Fenical W (2005) Marine actinomycete diversity and natural product discovery. *Antonie V. Leeuwenhoek* 87:43-48

Jiang C, Xu L. (1993) . Actinomycete diversity in unusual habitats. *Actinomycetes*; 4(2):47-57.

Jing, G., Sun, Z., Wang, H., Gong, Y., Huang,

S., Ning, K., Xu, J. and Su, X., (2017). Parallel-META 3: comprehensive taxonomical and functional analysis platform for efficient comparison of microbial communities. *Scientific Rep*., 7(1), pp. 1-11.

Johnson, D. B., Bacelar-Nicolau, P., Okibe, N., Thomas, A. and Hallberg, K. B. (2009). Ferrimicrobium acidiphilum gen. nov., sp. nov. and Ferrithrix thermotolerans gen. nov., sp. nov. : heterotrophic, iron-oxidizing, extremely acidophilic actinobacteria. *Int. J. Sys. Evol. Microbiol.*, 59(5), pp. 1082-1089.

Jones, S. E. and Elliot, M. A., (2017). Streptomyces exploration: competition, volatile communication and new bacterial behaviours. *Trends Microbiol*., 25(7), pp. 522-531.

Joutey, N. T., Bahafid, W., Sayel, H. and El Ghachtouli, N. (2013). Biodegradation: involved microorganisms and genetically engineered microorganisms. Biodegradation-life of Sci., pp. 289-320.

Jurado, V., Kroppenstedt, R. M., Saiz-Jimenez, C., Klenk, H. P., Mouniee, D., Laiz, L., Couble, A., Poetter, G., Boiron, P. and Rodriguez-Nava, V. (2009). Hoyosella altamirensis gen. nov., sp. nov., a new member of the order Actinomycetales isolated from a cave biofilm. *Int. J. Syst. Evol. Microbiol*., 59(12), pp. 3105-3110.

Kakirde, K. S., Parsley, L. C. and Liles, M. R., (2010.) Size does matter: application-driven approaches for soil metagenomics. *Soil Biol. and Biochem.,* 42(11), pp. 1911-1923.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M., (2012.) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids Res*., 40 (D1), pp. D109-D114.

Karmakar, D. K., 2005. The Tea Industry in India: A survey, dept of Economic analysis and Research. NABARD, Mumbai, India.

Kathiresan K, Bingham BL (2001). Biology of mangroves and mangrove ecosystems. *Adv. Mar. Biol.* 40; 81-251.

Kaur, T., Jasrotia, S., Ohri, P. and Manhas, R. K., (2016). Evaluation of in vitro and in vivo nematicidal potential of a multifunctional streptomycete, Streptomyces hydrogenans strain DH16 against Meloidogyne incognita. *Microbiological Res*., 192, pp. 247-252.

Kechin, A., Boyarskikh, U., Kel, A. and Filipenko, M., (2017). cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Computational Biol.*, 24(11), pp. 1138-1143.

Kennedy, J., O'leary, N. D., Kiran, G. S., Morrissey, J. P., O'Gara, F., Selvin, J. and Dobson, A. D. W., (2011). Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *J. Appl. Microbiol*., 111(4), pp. 787-799.

Kerfahi, D., Tripathi, B. M., Porazinska, D. L., Park, J., Go, R. and Adams, J. M., (2016). Do tropical rain forest soils have greater nematode diversity than High Arctic tundra? A metagenetic comparison of Malaysia and Svalbard. *Global Ecol. Biogeo*., 25(6), pp. 716-728.

Kesho, A., 2020. Microbial bio-pesticides and their use in integrated pest management. *Chem Biomol Eng*, *5*, p. 26.

Khattab, A. I. A. F. (2016). Isolation and Characterization of Antibiotic from Streptomyces sp. from Red Sea-Port Sudan (Doctoral dissertation, Sudan University of Sci. & Technology).

Kheiralla, Z. H., Hewedy, M. A., Mohammed, H. R. and Darwesh, O. M. (2016). Isolation of Pigment Producing Actinomycetes from Rhizosphere Soil and Application It in Textiles Dyeing. *Res. J. Pharma. Biolo. Chem. Sci.*

Kim S. B,Lonsdale J,Seong C. N,Goodfellow M (2003) Streptacidiphilus gen. nov., acidophilic actinomycetes with wall chemotype I and emendation of the family Streptomycetaceae (Waksman and Henrici (1943)AL) emend. Rainey et al. 1997. *Antonie Van Leeuwenhoek* 83:107–116. doi:10. 1023/A:1023397724023.

Kim T. K., Garson M. J, Fuerst J. A (2005). Marine actinobacteria related to the Salinospora group from the Great Barrier Reef sponge Pseudoceratina clavata. *Env. Microbiol*. 7:509-518

Kim, J. H., (2019). Gene Expression Data Analysis. In Genome Data Analysis (pp. 95-

120). Springer, Singapore.

Kim, M., Cho, A., Lim, H. S., Hong, S. G., Kim, J. H., Lee, J., Choi, T., Ahn, T. S. and Kim, O. S., (2015). Highly heterogeneous soil bacterial communities around Terra Nova Bay of northern Victoria Land, Antarctica. *PLoS One*, 10(3), p. e0119966.

Kinashi, H., Shimaji, M. and Sakai, A. (1987). Giant linear plasmids in Streptomyces which code for antibiotic biosynthesis genes. *Nature*, 328(6129), pp. 454-456.

Kirk, J. L., Beaudette, L. A., Hart, M., Moutoglis, P., Klironomos, J. N., Lee, H. and Trevors, J. T., (2004). Methods of studying soil microbial diversity. J. microbiological methods, 58(2), pp. 169-188.

Kladivko, E. J., (2001). Tillage systems and soil ecology. *Soil and Tillage Res*., 61(1-2), pp. 61-76.

Knight, R. D., Freeland, S. J. and Landweber, L. F., (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol*., 2(4), pp. 1-13.

Köberl, M., Müller, H., Ramadan, E. M. and Berg, G. (2011). Desert farming benefits from microbial potential in arid soils and promotes diversity and plant health. *Plos One*, 6(9), p. e24452.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. and Wilson, R. K., (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*., 22(3), pp. 568-576.

Kumar, P., Badal, P. S., Singh, N. P. and Singh, R. P., (2008). Tea industry in India: Problems and prospects. *Ind. J. Agricultural Economics*, 63(902-2016-67960).

Küster, E. and Williams, S. T., (1964). Selection of media for isolation of streptomycetes. *Nature*, 202(4935), pp. 928-929.

Kyte, J. and Doolittle, R. F., (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol*., 157(1), pp. 105-132.

L. Shivlata, Tulasi Satyanarayana (2015). Thermophilic and alkaliphilic Actinobacteria: biology and potential applications. Front. Microbiol. 2015; 6: 1014. doi: 10. 3389/fmicb. 2015. 01014

Ladenstein, R., and Ren, B. (2006). Protein disulfides and protein disulfide oxidoreductases in hyperthermophiles. Fed. Eur. Biochem. Soc. Lett. 273, 4170–4185. doi: 10. 1111/j. 1742-4658. 2006. 05421.

Lam, K. S., 2007. New aspects of natural products in drug discovery. *Trends in Microbiol., 15*(6), pp. 279-289.

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Thurber, R. L. V., Knight, R. and Beiko, R. G., (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnol.* 31(9), pp. 814-821.

Lapierre, P. and Gogarten, J. P., (2009). Estimating the size of the bacterial pan-genome. *Trends in genetics,* 25(3), pp. 107-110.

Law, J. W. F., Pusparajah, P., Ab Mutalib, N. S., Wong, S. H., Goh, B. H. and Lee, L. H., (2019). A review on mangrove actinobacterial diversity: the roles of Streptomyces and novel species discovery. *Trends in genetics.,* 2(1).

Le Roes-Hill, M. and Prins, A. (2016). Biotechnological potential of oxidative enzymes from Actinobacteria. Actinobacteria -Basics and Biotechnological Applications.http://dx.doi.org/10.5772/61321

Lechevalier, H. A, Lechevalier, M. P. (1970). A critical evaluation of the genera of aerobic actinobacteria. The Actinomycetales the Jena Int. Symposium on Taxonomy, September 1968, pp 393-405

Lechevalier, H. and Holbert, P. E., 1965. Electron microscopic observation of the sporangial structure of a strain of Actinoplanes. *J. Bact., 89*(1), pp. 217-222.

Lechevalier, H. A., Lechevalier, M. P. and Holbert, P. E., 1966. Electron microscopic observation of the sporangial structure of strains of Actinoplanaceae. *J. Bact*., 92(4),

pp. 1228-1235.

Lechevalier, H. A. and Lechevalier, M. P., 1967. Biology of actinomycetes. *Annual Rev.in Microbiol., 21*(1), pp. 71-100

Lee, L. H., Zainal, N., Azman, A. S., Eng, S. K., Goh, B. H., Yin, W. F., Ab Mutalib, N. S. and Chan, K. G., (2014). Diversity and antimicrobial activities of actinobacteria isolated from tropical mangrove sediments in Malaysia. *The Scientific World J*., 2014.

Lee-Cruz, L., Edwards, D. P., Tripathi, B. M. and Adams, J. M., (2013). Impact of logging and forest conversion to oil palm plantations on soil bacterial communities in Borneo. *Appl. Environ. Microbiol*., 79(23), pp. 7290-7297.

Leng, J., Xie, L., Zhu, R., Yang, S., Gou, X., Li, S. and Mao, H., (2011). Dominant bacterial communities in the rumen of Gayals (Bos frontalis), Yaks (Bos grunniens) and Yunnan Yellow Cattle (Bos taurs) revealed by denaturing gradient gel electrophoresis. *Mol. Biol. Rep.*, 38(8), pp. 4863-4872.

Lengeler,J. W.,Drews,G.,andSchlegel,H. G. (1999). Biology of the Prokaryotes. Stuttgart :Blackwell.

Li, H., (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303. 3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), pp. 2078-2079.

Li, X., Yuan, Y., Huang, Y., Liu, H. W., Bi, Z., Yuan, Y. and Yang, P. B., (2018). A novel method of simultaneous NH4+ and NO3− removal using Fe cycling as a catalyst: Feammox coupled with NAFO. *Sci. of the Total Environment,* 631, pp. 153-157.

Lithwick, G. and Margalit, H., (2005). Relative predicted protein levels of functionally associated proteins are conserved across organisms. *Nucleic acids Res.,* 33(3), pp. 1051-1057.

Liu, Y., (2020) . A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Communication and Signaling*, 18(1), pp. 1-9.

Lladó, S., López-Mondéjar, R. and Baldrian, P., (2017). Forest soil bacteria: diversity, involvement in ecosystem processes, and response to global change. *Microbiol. Mol. Biol. Reviews,* 81(2), pp. e00063-16.

Lombard, N., Prestat, E., van Elsas, J. D. and Simonet, P., (2011). Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. *FEMS Microbiol. Ecol*., 78(1), pp. 31-49.

Lu, S., Lepo, J. E., Song, H. X., Guan, C. Y. and Zhang, Z. H., (2018). Increased rice yield in long-term crop rotation regimes through improved soil structure, rhizosphere microbial communities, and nutrient bioavailability in paddy soil. Biol. and Fertility of Soils, 54(8), pp. 909-923.

Luedemann, G. M., (1968). Geodermatophilus, a new genus of the Dermatophilaceae (Actinomycetales). *J. Bact*., 96(5), pp. 1848-1858.

Maciejewska, M., Adam, D., Naômé, A., Martinet, L., Tenconi, E., Całusińska, M., Delfosse, P., Hanikenne, M., Baurain, D., Compère, P. and Carnol, M., (2017). Assessment of the potential role of Streptomyces in cave moonmilk formation. *Front. Microbiol*., 8, p. 1181.

Maciejewska, M., Pessi, I. S., Arguelles-Arias, A., Noirfalise, P., Luis, G., Ongena, M., Barton, H., Carnol, M. and Rigali, S. (2015). Streptomyces lunaelactis sp. nov., a novel ferroverdin A-producing Streptomyces species isolated from a moonmilk speleothem. *Antonie Van Leeuwenhoek*, 107 (2), pp. 519-531.

Magarvey N. A, Keller J. M, Bernan V, Dworkin M, Sherman D. H (2004). Isolation and characterization of novel marine-derived actinomycete taxa rich in bioactive metabolites. *Appl. Environ. Microbiol*. 70:7520-7529

Malik, A., Kim, Y. R., Jang, I. H., Hwang, S., Oh, D. C. and Kim, S. B. (2020). Genome-based analysis for the bioactive potential of Streptomyces yeochonensis CN732, an acidophilic filamentous soil actinobacterium. *BMC genomics*, 21(1), p. 118.

Manivasagan, P., Venkatesan, J., Sivakumar,

K. and Kim, S. K. (2014). Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiol. Res*., 169(4), pp. 262-278

Mann, J., (2001). Natural products as immunosuppressive agents. *Nat. Prod. Rep*., 18(4), pp. 417-430.

Margesin, R. and Miteva, V., (2011). Diversity and ecology of psychrophilic microorganisms. *Res. in Microbiol*., 162(3), pp. 346-361.

Margesin, R., Schinner, F., Marx, J. C. and Gerday, C. eds. (2008). Psychrophiles: from biodiversity to biotechnology (pp. 1-462). Berlin: Springer.

Maskey, R. P, Li F. C., Qin, S, Fiebig,H. H., Laatsch, H. (2003). Chandrananimycins AC: production of novel anticancer antibiotics from a marine Actinomadura sp. isolate M048 by variation of medium composition and growth conditions. *J. Antibiot* 56:622-629

Maurice, C. F., Haiser, H. J. and Turnbaugh, P. J., (2013). Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell*, 152(1-2), pp. 39-50.

Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R., (2005). The microbial pan-genome. Current opinion in genetics & development, 15(6), pp. 589-594.

Mevs, U., Stackebrandt, E., Schumann, P., Gallikowski, C. A. and Hirsch, P., (2000) Modestobacter multiseptatus gen. nov., sp. nov., a budding actinomycete from soils of the Asgard Range (Transantarctic Mountains). *Int. J. Systematic and Evolutionary Microbiol*., 50(1), pp. 337-346.

Mishra, S., Lin, Z., Pang, S., Zhang, W., Bhatt, P. and Chen, S., (2021). Recent advanced technologies for the characterization of xenobiotic-degrading microorganisms and microbial communities. . *Front. in Bioengineering and Biotechnology*, 9, p. 31.

Montero-Calasanz, M., Göker, M., Pötter, G., Rohde, M., Spröer, C., Schumann, P., Gorbu a, A. A. and Klenk, H. P. (2012). Geodermatophilus arenarius sp. nova, a xerophilic actinomycete isolated from Saharan desert sand in Chad. *Extremophiles*, 16(6), pp. 903-909.

Moore, R. E. and Townsend, S. D., (2019). Temporal development of the infant gut microbiome. *Open Biol*., 9(9), p. 190128.

Morneau-Vaillancourt, G., 2021. Les dimensions du retrait social à l'enfance: associations avec les difficultés interpersonnelles, étiologie génétique et précurseurs tempéramentaux.

Nafis, A., Raklami, A., Bechtaoui, N., El Khalloufi, F., El Alaoui, A., Glick, B. R., Hafidi, M., Kouisni, L., Ouhdouch, Y. and Hassani, L. (2019). Actinobacteria from extreme niches in morocco and their plant growth-promoting potentials. *Diversity*, 11 (8), p. 139.

Naine, S. J., Devi, C. S. and Mohanasrinivasan, V. (2015). Antimicrobial, Antioxidant and Cytotoxic Activity of MarineStreptomyces parvulus VITJS11 Crude Extract. *Braz Arch Biol Technol*, 58(2), pp. 198-207.

Nannipieri, P., Ascher, J., Ceccherini, M., Landi, L., Pietramellara, G. and Renella, G., (2003). Microbial diversity and soil functions. *Euro. J. soil Sci*., 54(4), pp. 655-670.

Nasr-Eldin, M., Messiha, N., Othman, B., Megahed, A. and Elhalag, K., (2019). Induction of potato systemic resistance against the potato virus Y (PVY NTN), using crude filtrates of Streptomyces spp. under greenhouse conditions. *Egyptian J. Biological Pest Control*, 29(1), pp. 1-11.

Naya, H., Romero, H., Carels, N., Zavala, A. and Musto, H., (2001). Translational selection shapes codon usage in the GC-rich genome of Chlamydomonas reinhardtii. *FEBS letters,* 501(2-3), pp. 127-130.

Neemisha (2020) Role of Soil Organisms in Maintaining Soil Health, Ecosystem Functioning, and Sustaining Agricultural Production. In: Giri B., Varma A. (eds) Soil Health. Soil Biol., vol 59. *Springer*, Cham. https://doi. org/10. 1007/978-3-030-44364-1_17

Nesme, J., Achouak, W., Agathos, S. N., Bailey, M., Baldrian, P., Brunel, D., Frostegård, Å., Heulin, T., Jansson, J. K., Jurkevitch, E. and Kruus, K. L., (2016). Back to the future of soil metagenomics. *Front. Microbiol*., 7, p. 73.

Newman DJ, Cragg GM (2007). Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod*. 70:461-477

Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K. and Epstein, S. S., (2010). Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. *Appl. Environ. Microbiol*., 76(8), pp. 2445-2450.

Normand, P. and Benson, D. R. (2012) Family IV. Geodermatophilaceae Normand 2006, 2277VP (Effective publication: Normand, orso, cournoyer, jeannin, chapelon, dawson, evtushenko and misra 1996, 8.). *Bergey's Manual of System. Bact*., 5, p. 528.

Novakova, R., Knirschova, R., Farkasovsky, M., Feckova, L., Rehakova, A., Mingyar, E., et al. (2013). The gene cluster aur1 form the angucycline antibiotic auricin is located on a large linear plasmid pSA3239 in Steptomyces aureofaciens CCM 3239. *FEMS Microbiol. Lett.* 342, 130–137. doi: 10. 1111/1574-6968. 12095

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y. and Stepanauskas, R., (2013). Assembling genomes and mini-metagenomes from highly chimeric reads. In Annual international conference on research in computational molecular Biol. (pp. 158-170). Springer, Berlin, Heidelberg.

Omura, S., (2008). Ivermectin: 25 years and still going strong. *Int. J. Antimicrob. Agents*, 31(2), pp. 91-98.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M. and Vonstein, V., (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*., 42(D1), pp. D206-D214.

Pascual, J. A., Garcia, C., Hernandez, T., Moreno, J. L. and Ros, M., (2000). Soil microbial activity as a biomarker of degradation and remediation processes. *Soil Biol. and Biochem*., 32(13), pp. 1877-1883.

Pearce, D. A., Newsham, K., Thorne, M., Calvo-Bado, L., Krsek, M., Laskaris, P., Hodson, A. and Wellington, E. M., (2012). Metagenomic analysis of a southern maritime Antarctic soil. *Front. Microbiol*., 3, p. 403.

Peden, J. F., (1999). Analysis of codon usage [thesis]. [Nottingham (United Kingdom)]: University of Nottingham. CodonW: Correspondence analysis of codon usage. PhD Thesis.

Picard, C., Fioramonti, J., Francois, A., Robinson, T., Neant, F. and Matuchansky, C., 2005. bifidobacteria as probiotic agents–physiological effects and clinical benefits. *Aliment Pharmacol Ther.*, 22(6), pp. 495-512.

Pieper, D. H., (2005). Aerobic degradation of polychlorinated biphenyls. *Appl. Microbiol. biotechnology*, 67(2), pp. 170-191.

Pikovskaya, R. I., 1948. Mobilization of phosphorus in soil in connection with vital activity of some microbial species. *Mikrobiologiya*, *17*, pp. 362-370.

Ponge, J. F., (2015). The soil as an ecosystem. *Biol. and fertility soils*, 51(6), pp. 645-648.

Price, M. N., Dehal, P. S. and Arkin, A. P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one*, *5*(3), p. e9490.

Promnuan, Y., Kudo, T. and Chantawannakul, P. (2009). Actinomycetes isolated from beehives in Thailand. *World J. Microbiol. Biotechnol*., 25(9), pp. 1685-1689.

Pruitt, K. D., Tatusova, T. and Maglott, D. R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids Res*., 35(suppl_1), pp. D61-D65.

R. Anandan, D. Dharumadurai, G. P. Manogaran (2016), An introduction to actinobacteria D. Dhanasekaran, Y. Jiang (Eds.), Actinobacteria-Basics and Biotechnological Applications, *Intechopen* pp. 3-37

Radolf, J. D., Deka, R. K., Anand, A., Šmajs, D., Norgard, M. V. and Yang, X. F., (2016). Treponema pallidum, the syphilis spirochete: making a living as a stealth pathogen. *Nature Rev. Microbiol*., 14(12), pp. 744-759.

Raffa, C. M. and Chiampo, F., (2021). Bioremediation of agricultural soils polluted

with pesticides: *a review. Bioengineering*, 8 (7), p. 92.

Rajbangshi, P. R. and Nambiar, D., (2020). "Who will stand up for us?" the social determinants of health of women tea plantation workers in India. *Int. J. for equity in health*, 19(1), pp. 1-10.

Rajput, S., Hense, S. and Thankappan, K. R., 2021. Healthcare utilisation: a mixed-method study among tea garden workers in Indian context. *J. Health Res.*

Ramnath, L., Sithole, B. and Govinden, R., (2017). Identification of lipolytic enzymes isolated from bacteria indigenous to Eucalyptus wood species for application in the pulping industry. *Biotechnol. Rep.*, 15, pp. 114-124.

Rancourt, M. and Lechevalier, H. A., (1963). Electron microscopic observation of the sporangial structure of an actinomycete, Microellobosporia flavea. *Microbiol*., 31(3), pp. 495-498.

Rashid, M. I., Mujawar, L. H., Shahzad, T., Almeelbi, T., Ismail, I. M. and Oves, M., (2016). Bacteria and fungi can contribute to nutrients bioavailability and aggregate formation in degraded soils. *Microbiological Res*., 183, pp. 26-41.

Rateb, M. E., Houssen, W. E., Harrison, W. T., Deng, H., Okoro, C. K., Asenjo, J. A., Andrews, B. A., Bull, A. T., Goodfellow, M., Ebel, R. and Jaspars, M. (2011). Diverse metabolic profiles of a Streptomyces strain isolated from a hyper-arid environment. *J. Nat. Prod*. 74(9), pp. 1965-1971.

Ravi, L. and Kannabiran, K., (2018). Extraction and identification of gancidin W from marine streptomyces sp. VITLGK012. *Ind. J. Pharma. Sci*., 80(6), pp. 1093-1099.

Raynaud, X. and Nunan, N., (2014). Spatial ecology of bacteria at the microscale in soil. *PloS one*, 9(1), p. e87217.

Rego, A., Raio, F., Martins, T. P., Ribeiro, H., Sousa, A. G., Séneca, J., Baptista, M. S., Lee, C. K., Cary, S. C., Ramos, V. and Carvalho, M. F., (2019.) Actinobacteria and cyanobacteria diversity in terrestrial antarctic microenvironments evaluated by culture-dependent and independent methods. *Front. Microbiol.,* 10, p. 1018.

Rice, P., Longden, I. and Bleasby, A., (2000). EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, 16 (6), pp. 276-277.

Rice, E. L. and Pancholy, S. K., 1972. Inhibition of nitrification by climax ecosystems. *American J. Botany*, *59*(10), pp. 1033-1040.

Rodríguez, A. N. I., (2010). Algunos elementos para comprender la dinámica de trabajo en la UPN, un caso, la Unidad 094 DF Centro (Doctoral dissertation, UPN-94).

Roes-Hill, L., Durrell, K., Prins, A. and Meyers, P. R., 2018. Streptosporangium minutum sp. nov., isolated from garden soil exposed to microwave radiation. *The J. Antibiotics*, *71*(6), pp. 564-574.

Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., Daroub, S. H., Camargo, F. A., Farmerie, W. G. and Triplett, E. W., (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J*., 1(4), pp. 283-290.

Roymondal, U., Das, S. and Sahoo, S., (2009) Predicting gene expression level from relative codon usage bias: an application to Escherichia coli genome. *DNA Res*., 16(1), pp. 13-30.

Ruiz, L. M., Armengol, G., Habeych, E. and Orduz, S., (2006). A theoretical analysis of codon adaptation index of the Boophilus microplus bm86 gene directed to the optimization of a DNA vaccine. *J. theoretical Biol*., 239(4), pp. 445-449.

Saddhe, A. A., Jamdade, R. A. and Kumar, K. (2016). Assessment of mangroves from Goa, west coast India using DNA barcode. *SpringerPlus,* 5(1), p. 1554.

Saha, S., Yauvana, V., Chakraborty, S. and Sanyal, D., (2019). Synthesis and characterization of polyvinylidene-fluoride (PVDF) nanofiber for application as piezoelectric force sensor. Materials Today: Proceedings, 18, pp. 1450-1458.

Sahil, K., Prashant, B., Akanksha, M., Premjeet, S. and Devashish, R.,(2011). Gas chromatography-mass spectrometry: applications. *Int. J. pharmaceutical & biological archives*, 2(6), pp. 1544-1560.

Sakaguchi, K. (1990). Invertrons, a class of structurally and functionally related genetic elements that include linear DNA plasmids, transposable elements and genomes of adeno -type viruses. *Microbiol. Rev.* 54, 66–74.

Samavat S, Samavat S, Mafakheri S, Shakouri MJ (2012) Promoting Common Bean Growth and Nitrogen Fixation by the Co-Inoculation of Rhizobium and Pseudomonas fluorescens Isolates. *Bulgarian J. Agricultural Sci*. 18:387-395

Saikia, A. K., (2019) Problems and prospects of small tea growers of Assam. PhD Thesis, Rajiv Gandhi University.

Santos, A., Núñez-Montero, K., Lamilla, C., Pavez, M., Quezada-Solís, D. and Barrientos, L., (2020). Antifungal activity screening of antarctic actinobacteria against phytopathogenic fungi. Acta Biol. Colomb., 25(2), pp. 353-358.

Sapkota, A., Thapa, A., Budhathoki, A., Sainju, M., Shrestha, P. and Aryal, S., (2020). Isolation, characterization, and screening of antimicrobial-producing actinomycetes from soil samples. *Int. J. Microbiol*., 2020.

Sarkar, I., Kar, P., Sen, G., Chhetri, S., Bhattacharya, M., Bhattacharyya, S. and Sen, A., (2022). Metagenomic outlooks of microbial dynamics influenced by organic manure in tea garden soils of North Bengal, India. *Archives Microbiol*., 204(1), pp. 1-9.

Sarkar, I., Tisa, L. S., Gtari, M. and Sen, A., (2018). Biosynthetic energy cost of potentially highly expressed proteins vary with niche in selected actinobacteria. *J. Basic Microbiol*., 58(2), pp. 154-161.

Sarkar, S., 2019. Labour Migration in the Tea Plantations: Colonial and Neo-Liberal Trajectories of Plantation Labour in the Dooars Tea Belt of West Bengal. *J. Migration Affairs*, *2*(1), pp. 25-43.

Schaeffer, A., Nannipieri, P., Kästner, M., Schmidt, B. and Botterweck, J., (2015). From humic substances to soil organic matter–microbial contributions. In honour of Konrad Haider and James P. Martin for their outstanding research contribution to soil science. *J. Soils and Sediments*, 15(9), pp. 1865-1881.

Schmidt, H. & Eickhorst, T. (2013). Spatio-temporal variability of microbial abundance and community structure in the puddled layer of a paddy soil cultivated with wetland rice (Oryza sativa L.). *Appl. Soil Ecol.* 72, 93–102.

Schneider K, Nicholson G, Strobele M, Baur S, Niehaus J, Fiedler HP, Sussmuth R. D(2006) The structures of fluostatins C, D and E, novel members of the fluostatin family. *J Antibiot*. 59(2):105.

Seemann, T., (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), pp. 2068-2069.

Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P (2014) Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders â€˜Frankialesâ€™ and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov. Geodermatophilales ord. nov. Acidothermales ord. nov. and Nakamurellales ord. nov. *Int. J. Syst. and Evol. Microbiol*. 64:3821-3832

Sen, G., Sur, S., Bose, D., Mondal, U., Furnholm, T., Bothra, A., Tisa, L. and Sen, A., (2007). Analysis of codon usage patterns and predicted highly expressed genes for six phytopathogenic Xanthomonas genomes shows a high degree of conservation. *In silico Biol*., 7(4, 5), pp. 547-558.

Ser, H. L., Ab Mutalib, N. S., Yin, W. F., Chan, K. G., Goh, B. H. and Lee, L. H. (2015) . Evaluation of antioxidative and cytotoxic activities of Streptomyces pluripotens MUSC 137 isolated from mangrove soil in Malaysia. *Front. Microbiol*., 6, p. 1398.

Sghaier, H., Hezbri, K., Ghodhbane-Gtari, F., Pujic, P., Sen, A., Daffonchio, D., Boudabous, A., Tisa, L. S., Klenk, H. P., Armengaud, J. and Normand, P. (2016). Stone-dwelling actinobacteria Blastococcus saxobsidens, Modestobacter marinus and Geodermatophilus obscurus proteogenomes. *ISME J*., tom10(1), pp. 21-29.

Shan LI, Tao CH, Jin-song ZH, Rong-biao XI, Rong-gui HU, Shui-qing ZH, Mi-lan WA. (2014) Characteristics of soil organic carbon mineralization at different temperatures in paddy soils under long-term fertilization.

Yingyong Shengtai Xuebao1;25(5).

Sharma, P. and Thakur, D., 2020. Antimicrobial biosynthetic potential and diversity of culturable soil actinobacteria from forest ecosystems of Northeast India. *Scientific Rep*., 10(1), pp. 1-18.

Sharp, P. M. and Li, W. H., (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. evolution,* 24(1), pp. 28-38.

Sharp, P. M. and Li, W. H., (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids Res*., 15 (3), pp. 1281-1295.

Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F. and Sockett, R. E., (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic acids Res*., 33(4), pp. 1141-1153.

Shen, Z., Gan, Z., Zhang, F., Yi, X., Zhang, J. and Wan, X., (2020). Analysis of codon usage patterns in citrus based on coding sequence data. *BMC genomics*, 21(5), pp. 1-10.

Shepherd, M. D., Kharel, M. K., Bosserman, M. A. and Rohr, J., (2010). Laboratory maintenance of Streptomyces species. *Cur. prot. in Microbiol*., 18(1), pp. 10E-1.

Sherman, R. M. and Salzberg, S. L., 2020. Pan -genomics in the human genome era. *Nature Rev.Genetics,* 21(4), pp. 243-254.

Shin, S. K., Cho, Y. J. and Yi, H. (2017). Synonymy of Micropolyspora internatus and Saccharomonospora viridis and Emended Description of Saccharomonospora viridis. *J. Microbiol. Biotechnol.* 27(1), pp. 149-154.

Shirling E. B., Gottlieb D. (1966); Methods for characterisation of Streptomyces species. *Int. J. Systematic Bact*. 16:313–340

Shivlata, L. and Satyanarayana, T. (2015). Thermophilic and alkaliphilic Actinobacteria: Biol. and potential applications. *Front Microbiol* 6: 1014. Go to original source.

Shrestha, G. and Thapa, R. B., (2015). Tea pests and pesticide problems and integrated management. *J. Agric. and Environ*. 16, pp. 188-200.

Sikorski, J., (2015). The prokaryotic Biol. of soil. *Soil organ.*, 87(1), pp. 1-28.

Singh, R. and Dubey, A. K. (2018). Diversity and applications of endophytic actinobacteria of plants in special and other ecological niches. *Front. Microbiol*., 9, p. 1767.

Singh, S. P., Purohit, M. K., Aoyagi, C., Kitaoka, M., and Hayashi, K. (2010). Effect of growth temperature, induction and molecular chaperones on the solubilization of over-expressed cellobiose phosphorylase from Cellvibrio gilvus under in vivo conditions. Biotechnol. Bioprocess. Eng. 15, 273–276. doi: 10. 1007/s12257-009-0023-1 Singh, S. P., Shukla, R. J., and K

Singh, S., Verma, E. and Mishra, A. K., 2020. In silico molecular docking analysis of cancer biomarkers with GC/MS identified compounds of Scytonema sp. Network Modeling Analysis in Health Informatics and Bioinformatics, 9(1), pp. 1-14.

Sivakumar, K. (2001). Actinobacteria of an Indian mangrove (Pichavaram) environment: An inventory. Annamalai University, India:91 -94

Sivasankar, P., Priyanka, K., Rekadwad, B., Sivakumar, K., Thangaradjou, T., Poongodi, S., Manimurali, R., Bhaskar, P. V. and Anilkumar, N. (2018). Actinobacterial community structure in the Polar Frontal waters of the Southern Ocean of the Antarctica using Geographic Information System (GIS): A novel approach to study Ocean Microbiome. Data in brief, 17, pp. 1307-1313.

Smith, A. J, Hall, V., Thakker, B., Gemmell, C. G. (2005). Antimicrobial susceptibility testing of Actinomyces species with 12 antimicrobial agents. J. Antimicrob. Chemother. Aug;56 (2):407-9. doi: 10. 1093/jac/dki206.

Solanki, R,, Khanna, M., Lal, R. (2008). Bioactive compounds from marine actinobacteria. IJAM 48:410-431

Stach, J. E, Bull, A. T. (2005). Estimating and comparing the diversity of marine actinobacteria. Antonie Van Leeuwenhoek. Jan; 87(1):3-9. doi: 10. 1007/s10482-004-6524-1. PMID: 15726285

Stach, J. E. M., Maldonado, L. A., Ward, A. C., Bull, A. T., Goodfellow, M. (2004). Williamsiamaris sp. nov. a novel actinomycete isolated from the Sea of Japan. Int. J. Syst. *Evol. Microbiol*. 54:191-194

Subhash, G. S., Karad, D. D. and Kulkarni, S. W. (2016). Screening of Tyrosinase Producing Soil Actinomycetes from Shirala Region of Maharashtra, India. Int. *J. Curr. Microbiol. App. Sci,* 5(3), pp. 345-353.

Suela Silva, M., Naves Sales, A., Teixeira Magalhães-Guedes, K., Ribeiro Dias, D. and Schwan, R. F., (2013). Brazilian Cerrado soil actinobacteria ecology. *BioMed Res. Int*., 2013.

Suihko, M. L., Kroppenstedt, R. M., and Stackebrandt, E. (2006). Occurrenceand characterizationofactinobacteriaandthermoactinomycetesisolatedfrompulp and board samples containing recycled fibres. *J. Ind. Microbiol. Biotechnol.* 33, 183–191. doi:10. 1007/s10295-005-005

Takizawa, M., Colwell, R. R., Hill, R. T. (1993). Isolation and diversity of actinobacteria in the Chesapeake Bay. *Appl Environ Microbiol* 59:997-1002

Tamura, K., Dudley, J., Nei, M. and Kumar, S., (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4. 0. Mol. Biol. and evolution, 24(8), pp. 1596-1599.

Tan H, Deng Z, Cao L (2009) Isolation and characterization of actinomycetes from healthy goat faeces. Lett *Appl Microbiol*. 49 (2):248-253.

Tettelin, H., Riley, D., Cattuto, C. and Medini, D., (2008). Comparative genomics: the bacterial pan-genome. *Cur. Opin. Microbiol*., 11(5), pp. 472-477.

Thumar, J. T., Dhulia, K. and Singh, S. P., (2010). Isolation and partial purification of an antimicrobial agent from halotolerant alkaliphilic Streptomyces aburaviensis strain Kut-8. *World J. Microbiol. Biotechnol*. 26 (11), pp. 2081-2087.

Tidjani, A. R., Lorenzi, J. N., Toussaint, M., van Dijk, E., Naquin, D., Lespinet, O., Bontemps, C. and Leblond, P., (2019). Genome sequences of 11 conspecific Streptomyces sp. strains. *Microbiol. Resource Announce.,* 8(38), pp. e00863-19.

Tischler, D., van Berkel, W. J. H., Fraaije, M. W., eds. (2019). Actinobacteria, a Source of Biocatalytic Tools. Lausanne: . Front. Media. doi: 10. 3389/978-2-88945-922-3

Tiwari, K. and Gupta, R. K., (2013). Diversity and isolation of rare actinomycetes: an overview. *Cur. Rev. Microbiol*., 39(3), pp. 256-294.

Tomlinson, P. B., (2016). The botany of mangroves. Cambridge University Press.

Torsvik, V., Goksøyr, J. and Daae, F. L., (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol*., 56(3), pp. 782-787.

Tortora, G. J., Funke, B. R., and Case, C. L. (2007) Microbiogy: An Introduction. SanFrancisco, CA: Pearson Benjamin Cummings.

Trenozhnikova, L. and Azizan, A., (2018). Discovery of actinomycetes from extreme environments with potential to produce novel antibiotics. CAJGH, 7(1).

Tripathi, B. M., Kim, M., Lai-Hoe, A., Shukor, N. A., Rahim, R. A., Go, R. and Adams, J. M., 2013. pH dominates variation in tropical soil archaeal diversity and community structure. *FEMS Microbiol. Ecol.,* 86(2), pp. 303-311.

Tripathi, B. M., Kim, M., Singh, D., Lee-Cruz, L., Lai-Hoe, A., Ainuddin, A. N., Go, R., Rahim, R. A., Husni, M. H. A., Chun, J. and Adams, J. M., (2012). Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microbial Ecol*., 64 (2), pp. 474-484.

Tsai, C. T., Lin, C. H. and Chang, C. Y., (2007). Analysis of codon usage bias and base compositional constraints in iridovirus genomes. *Virus Res*., 126(1-2), pp. 196-206.

Van Teeseling, M. C., Pol, A., Harhangi, H. R., van der Zwart, S., Jetten, M. S., Op den Camp, H. J. and van Niftrik, L., (2014). Expanding the verrucomicrobial methanotrophic world: description of three novel species of Methylacidimicrobium gen. nov. *Appl. Environ. Microbiol*., 80(21), pp. 6782-6791.

Velho-Pereira, S. and Kamat, N. M., (2011). Antimicrobial screening of actinobacteria using a modified cross-streak method. Ind. *J. pharmaceutical Sci.*, 73(2), p. 223.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F. and van Sinderen, D., (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.,* 71(3), pp. 495-548.

Waksman SA, Curtis RE (1918) The occurrence of actinobacteria in the soil. Soil *Sci* 6:309-319

Vesth, T., Lagesen, K., Acar, Ö. and Ussery, D., (2013). CMG-biotools, a free workbench for basic comparative microbial genomics. *PloS one*, 8(4), p. e60120.

Wade, W., (2002). Unculturable bacteria—the uncharacterized organisms that cause oral infections. *J. the Royal Society of Medicine*, 95(2), pp. 81-83.

*Waksman, S. A., 1961. The Actinomycetes. Vol. II. Classification, identification and descriptions of genera and species. *The Actinomycetes. Vol. II. Classification, identification and descriptions of genera and species.*

Waksman, S. A., 1957. Species concept among the actinomycetes with special reference to the genus Streptomyces. *Bacteriol. Rev., 21*(1), pp. 1-29.

Waldrop, M. P., Balser, T. C. and Firestone, M. K., (2000). Linking microbial community composition to function in a tropical soil. *Soil Biol. and Biochem.*, 32(13), pp. 1837-1846.

Walkley, A. and Black, I. A., (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci*., 37(1), pp. 29-38.

Wang, L., Li, Y. and Li, Y., 2019. Metal ions driven production, characterization and bioactivity of extracellular melanin from Streptomyces sp. ZL-24. *Int. J. biological macromol. 123*, pp. 521-530.

Wang, H., Liu, N., Xi, L., Rong, X., Ruan, J. and Huang, Y., (2011). Genetic screening strategy for rapid access to polyether ionophore producers and products in actinomycetes. *Appl. Environ Microbiol*, 77 (10), pp. 3433-3442.

Watve MG, Shete AM, Jadhav N, Wagh SA, Shelar SP, Chakraborti SS, Botre AP, Kulkarni AA (1999) Myxobacterial diversity of Indian soils "How many species do we have? *Current Sci*. :1089-1095

Willerslev, E., Hansen, A. J., Rønn, R., Brand, T. B., Barnes, I., Wiuf, C., Gilichinsky, D., Mitchell, D. and Cooper, A., (2004). Long-term persistence of bacterial DNA. *Current Biol.*, 14(1), pp. R9-R10.

*William, S., Feil, H. and Copeland, A., (2012). Bacterial genomic DNA isolation using CTAB. Sigma, 50(6876). Protocol Book.

. Xu, J., Li, X., Shi, Z., Li, R. and Li, Q., 2018. Bacterial carbon cycling in the river plume in the northern South China Sea during summer. *J. Geophysical Res. : Oceans*, *123*(11), pp. 8106-8121.

Xu J, Gordon JI (2003) Honor thy symbionts. *Proc. Natl. Acad. Sci*. 100:10452-10459

Witzgall, K., Vidal, A., Schubert, D. I., Höschen, C., Schweizer, S. A., Buegger, F., Pouteau, V., Chenu, C. and Mueller, C. W., (2021). Particulate organic matter as a functional soil component for persistent soil organic carbon. *Nature Comm.*,12(1), pp. 1-10.

Wong, E. H., Smith, D. K., Rabadan, R., Peiris, M. and Poon, L. L., (2010). Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. *BMC evolutionary Biol*., 10(1), pp. 1-14.

Wood, D. E. and Salzberg, S. L., (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*., 15(3), pp. 1-12.

Wright, F., (1990). The 'effective number of codons' used in a gene. *Gene*, 87(1), pp. 23-29.

Wu, G., Culley, D. E. and Zhang, W., (2005). Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. *Microbiol*., 151(7), pp. 2175-2187.

Wu, T. D. and Watanabe, C. K., (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9), pp. 1859-1875.

Xia, X. and Xie, Z., (2001). DAMBE: software package for data analysis in molecular Biol. and evolution. *J. Heredity*, 92(4), pp. 371-373.

Xia, X., (2017). DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution. J. *Heredity*, 108(4), pp. 431-437.

Xu, Z., Masuda, Y., Itoh, H., Ushijima, N., Shiratori, Y. and Senoo, K. (2019). Geomonas oryzae gen. nov., sp. nova, Geomonas edaphica sp. nov., Geomonas ferrireducens sp. nov., Geomonas terrae sp. nov., Four Ferric-Reducing Bacteria Isolated From Paddy Soil, and Reclassification of Three Species of the Genus Geobacter as Members of the Genus Geomonas gen. nov. *Front. Microbiol*., 10, p. 2201.

Xue, P. P., Carrillo, Y., Pino, V., Minasny, B. and McBratney, A. B.,(2018). Soil properties drive microbial community structure in a large scale transect in South Eastern Australia. *Scientific Rep*., 8(1), pp. 1-11.

Yandigeri, M. S., (2021). Role of Actinomycetes in Insect Pest and Plant Disease Management. Biopesticides in Horticultural Crops, pp. 152-172.

Yang, J. C., Huang, J. H., Pan, Q. M., Tang, J. W. and Han, X. G., (2004). Long-term impacts of land-use change on dynamics of tropical soil carbon and nitrogen pools. J. *Environmental Sci*., 16(2), pp. 256-261.

Yoon, S. H., Ha, S. M., Lim, J., Kwon, S. and Chun, J., (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek*, 110(10), pp. 1281-1286.

Yu, Q., Hu, X., Ma, J., Ye, J., Sun, W., Wang, Q. and Lin, H., (2020). Effects of long-term organic material applications on soil carbon and nitrogen fractions in paddy fields. Soil and *Tillage Res*., 196, p. 104483.

Zarilla, K. A. and Perry, J. J. (1986) . Thermoleophilum album gen. nov., sp. nov. Validation of the Publication of New Names and New Combinations Previously Effectively Published Outside the IJSB, List, (20), pp. 354-356.

Zhang, G., Ma, X., Niu, F., Dong, M., Feng, H., An, L. and Cheng, G. (2007). Diversity and distribution of alkaliphilic psychrotolerant bacteria in the Qinghai–Tibet Plateau permafrost region. *Extremophiles*, 11 (3), pp. 415-424.

Zhang, M., Dong, L., Wang, Y., Bai, X., Ma, Z., Yu, X. and Zhao, Z., (2021). The response of soil microbial communities to soil erodibility depends on the plant and soil properties in semiarid regions. Land Degradation & Development, 32(11), pp. 3180-3193.

Zhang, M., Fu, X. H., Fang, W. T. and Zou, X., 2007. Soil organic carbon in pure rubber and tea-rubber plantations in southwestern China. *Tropical Ecol*., 48(2), p. 201.

Zhang, Q. C., Shamsi, I. H., Xu, D. T., Wang, G. H., Lin, X. Y., Jilani, G., Hussain, N. and Chaudhry, A. N., 2012. Chemical fertilizer and organic manure inputs in soil exhibit a vice versa pattern of microbial community structure. *Appl. Soil Ecology*, *57*, pp. 1-8.

Zhou, M. and Li, X., (2009). Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. *Mol. Biol. Rep*., 36(8), pp. 2039-2046.

Zhou, Z., Wang, C., Zheng, M., Jiang, L. and Luo, Y., (2017). Patterns and mechanisms of responses by soil microbial communities to nitrogen addition. *Soil Biol. and Biochem*., 115, pp. 433-441.

Zhu, P., Li, Q., Wang, G. (2008). Unique microbial signatures of the alien Hawaiian marine sponge Suberites zeteki. *Microb. Ecol.* 55:406-414

Zitouni, A., Lamari, L., Boudjella, H., Badji, B., Sabaou, N., Gaouar, A., Mathieu, F., Lebrihi, A. and Labeda, D. P., (2004). Saccharothrix algeriensis sp. nov., isolated from Saharan soil. *Int. J. system. Evol. Microbiol.*, 54(4), pp. 1377-1381.■

Appendix-A

List of Publications

Sen G, I Sarkar, S Chetrri, P Kar, A Roy, A Sen, M Bhattacharya (2022). Rhizospheric soil metabarcoding analysis of *Alnus nepalensis* from Darjeeling hills reveals the abundance of nitrogen-fixing symbiotic microbes. *J Forest Res.* DOI: 10.1080/13416979.2022.2037813 **IF: 1.269.**

Sen G, S Sur, D Bose, U Mondal, T Furnholm, A Bothra, L Tisa and A Sen (2007) Analysis of codon usage patterns and predicted highly expressed genes for six phytopathogenic Xanthomonas genomes shows a high degree of conservation. *In silico Biol.* 7:0039 http:// www.bioinfo.de/isb/2007/07/0039/

Sarkar I, **G Sen** and **A Sen** (2020) Methods for whole-genome analysis of Actinobacteria through Bioinformatics approaches. In Methods in Actinobacteriology, Springer Protocols Handbooks (Springer Protocols), Springer Nature, New York.

Chhettri S, **G Sen** and A Sen (2022). Isolation and identification of Streptomyces sp. Producing agroactive enzymes with biocontrol potential. *Ind J Applied and Pure Bio*. 140 -146

Chhettri S, **G Sen** and A Sen (2022). Morphology, diversity and importance of Actinobacteria. *In* 'Biodiversity and Sustainable Resource Management (Basic to Research)' Ed.: D Das. Bharti Publications

Sarkar I, **G Sen**, M Bhattyacharya, S Bhattacharyya and **A Sen** (2021) *In silico* inquest reveals the efficacy of Cannabis in the treatment of Post-Covid-19 related neuro-degeneration. *J. Biomolecul. Struct. Dyna.*. https://doi.org/10.1080/07391102.2021. 1905556. **IF: 3.31**

Sur S, **G Sen**, S Thakur, AK Bothra and **A Sen** (2009) *In silico* analysis of evolution in swine flu viral genomes through re-assortment by promulgation and mutation. *Biotechnol*. 8:434-441.

Sarkar I, P Kar, **G Sen**, S. Chhetri, M Bhattacharya, S Bhattacharyya, **A Sen** (2021). Metagenomic outlooks of microbial dynamics influenced by organic manure in tea garden soils of North Bengal, India. *Archives Microbiolo.* DOI: 10.1007/s00203-021-02635-6. **IF: 2.552**

Ramachandran, P., et al. (2000). Evidence for association of a viroid with tapping panel dryness syndrome of rubber (*Hevea Brasiliense*). *Plant disease*, 84(10), pp.1155-1155.**IF: 4.438**

Appendix-B

Genome Projects

Whole Genomes

Sen,G., Sen,A., Chhettri,S., Sarkar,I. and Bhattacharya,M. Whole genome sequencing of *Streptomyces* sp. Tea02. https://www.ncbi.nlm.nih.gov/nuccore/JAIGNW000000000.1

Sen,G. and Sen,A. Whole genome sequencing of *Streptomyces actuosus* VRA1 https://www.ncbi.nlm.nih.gov/nuccore/NZ_JAFFZS000000000.1

Sen,G., Sen,A., Chhettri,S., Sarkar,I. and Bhattacharya,M. Whole genome sequencing of *Streptomyces* sp. Tea10. https://www.ncbi.nlm.nih.gov/nuccore/jaignv000000000.1

Sen,G., Bhattacharya,M., Sarkar,I., Chettri,S. and Sen,A.Whole genome sequencing of *Streptomyces* species from India. https://www.ncbi.nlm.nih.gov/nuccore/NZ_JAFVLN000000000.1

Partial Genomes

VRA-1	MW332556.1	Streptomyces actuosus
VRA-3	MW585686.1	Streptomyces sp. CT9210B3
VRA-9	MW585687.1	Streptomyces tanashiensis
VRA-10		Streptomyces
VRA-12	MW332566.1	Streptomyces
VRA-14	MW332567.1	Streptomyces
VRA-16	MW332568.1	Streptomyces sp.
VRA-17	MW332569.1	Streptomyces sp.
VRA-19	OL 851820.1	Streptomyces
VR-05		Streptomyces
MTA.1	MW332570.1	Streptomyces
MTA.13	MW332571.1	Streptomyces
Tea02	MK299993.1	Streptomyces
TEA10	MK290326.1	Streptomyces
ARA.1	MW585688.1	Streptomyces fimbriatus
ARA.2		Streptomyces microflavus
ARA.3	MW585689.1	Streptomyces sp. PDP
ARA.4	MW585690.1	Streptomyces viridifaciens
KL02		Streptomyces

Genome Projects
Meta-sequence Submission

Sen,G., Comparative 16s metabarcoding of Alnus rhizosphere and nonrhizosphere. https://www.ncbi.nlm.nih.gov/biosample/SAMN25729411

Sen,G., Comparative 16s metabarcoding of Alnus rhizosphere and nonrhizosphere. https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN25729412

Sen,G., Comparative 16s metabarcoding of Alnus rhizosphere and nonrhizosphere. https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN25729413

Sen,G., Comparative 16s metabarcoding of Alnus rhizosphere and nonrhizosphere. https://www.ncbi.nlm.nih.gov/biosample/SAMN25729414

Sen,G., Comparative 16s metabarcoding of Alnus rhizosphere and nonrhizosphere. https://www.ncbi.nlm.nih.gov/biosample/SAMN25729415

Appendix-C

Buffers & Chemicals

CTAB- buffer

100mM Trizma Base (Sigma, Cat# T1503) (pH-8.0)

20mM EDTA (Merck India, Cat# 60841801001730) (pH-8.0)

1.4 M NaCl (Merck India, Cat#60640405001730)

2% (w/v) CTAB (Hexadecyl cetyl trimethyl ammonium bromide) (Sigma, Cat# H6269)

12.11g of molecular grade Trizma base was dissolved in 400 ml double distilled water, pH was adjusted to 8.0 and was divided into two parts of equal volume. To one part 7.44g EDTA was added and to the other part 81.8g NaCl and 20g CTAB. Both the parts were than mixed and the final volume was made up to 1000ml with double distilled water prior to autoclaving. The buffer was autoclaved at 121ºC and 15 psi for 20 mins and stored at room temperature for further use.

Note: Add 1% PVP (Polyvinylpyrrolidone) (Sigma, Cat #P5288) and 0.3% β-mercaptoethanol (Sigma, Cat# M3148) just before use.

5X TBE (Tris-borate-EDTA) buffer

Trizma base (Sigma, Cat# T1503) = 27 gm

Boric acid (Sigma, Cat# 15663)= 13.75 gm

0.5M EDTA (pH 8.0)=1.86 gm

All the reagents were dissolved separately and finally mixed together and the final volume was made up to 1000ml with double distilled water prior to autoclaving. The buffer was autoclaved at 121ºC and 15 psi for 20 mins and stored at room temperature for further use.

1X TE:

Tris- Cl (pH 8.0) (i.e. 10Mm) =0.6055gm

EDTA (pH 8.0) (i.e. 1mM) =0.186 gm

Both the reagents were dissolved separately and finally mixed together and the final volume was made up to 1000ml with double distilled water prior to autoclaving. The buffer was autoclaved at 121ºC and 15 psi for 20 mins and stored at room temperature for further use.

3M Sodium Acetate (Sigma, Cat# S9513):

The required amount of sodium acetate i.e.12.31 g was dissolved in 50ml double distilled water prior to autoclaving. The solution was autoclaved at 121ºC and 15 psi for 20 mins and stored at room temperature for further use.

6X gel loading buffer:

TYPE 3:

0.25% Bromophenol blue (Sigma, Cat# B0126)

0.25% Xylene cyanol FF (Sigma, Cat# X4126)

30% Glycerol (Merck India, Cat#61756005001730) in water

Store at 4ºC.

RNase A:

The RNase A enzyme (Sigma, Cat# R4875) was dissolved at a concentration of 10mg/ml in 0.01M sodium acetate (Sigma, Cat# S9513) (pH 5.2). The solution was heated at 100ºC for 15 minutes in a water bath and allowed to cool slowly to room temperature. The pH was adjusted by adding 1/10 volume of 1M Tris- Cl (pH 7.4) and stored at -20ºC for further use.

Note: Both 0.01M sodium acetate and 1M Tris-Cl were prepared and autoclaved at 121ºC and 15 psi for 20 mins prior to use.

The other chemical used for the molecular work are:

Ready Mix TM Taq PCR Reaction (Sigma, Cat# P4600)

Proteinase K (Sigma, USA. Cat #P2308

Chloroform (Merck India, Cat #822265):

Isoamyl alcohol (Merck India Cat# 8.18969.1000)

Phenol (Sigma, Cat #P4557)

Isopropanol (Merck India, Cat#17813)

Absolute ethyl alcohol (BDH, Cat#10107)

Agarose (Sigma, Cat#A9539)

Ethidium bromide (10mg/ml) (Hi Media, Cat# RM813)

Lambda DNA/ EcoRI/ HindIII double digest (Promega, Cat# PR-G1731)

100 bp ladder (Sigma, Cat#1473)

Culture Media

ISP 4 medium (HI-MEDIA) M359

Ingredients Gms / Litre

Starch, soluble 10.000

Dipotassium hydrogen phosphate 1.000

Magnesium sulphate heptahydrate 1.000

Sodium chloride 1.000

Ammonium sulphate 2.000

Calcium carbonate 2.000

Ferrous sulphate heptahydrate 0.001

Manganous chloride, heptahydrate 0.001

Zinc sulphate heptahydrate 0.001

Agar 20.000

Final pH (at 25°C) 7.2±0.2

Streptomyces Agar (HI-MEDIA) M1352

Ingredients Gms / Litre Malt extract 10.000 Yeast extract 4.000 Dextrose 4.000 Calcium carbonate 2.000 Agar 12.000

Actinomycetes isolation Agar(HI-MEDIA) M490

Sodium caseinate 2.000 L-Asparagine 0.100 Sodium propionate 4.000 Dipotassium phosphate 0.500 Magnesium sulphate 0.100 Ferrous sulphate 0.001 Agar 15.000 Final pH (at 25°C) 8.1±0.2

Nutrient Agar (HI-MEDIA) MM012

Peptone 10.000 Meat extract B # 10.000 Sodium chloride 5.000 Agar 12.000 pH after sterilization 7

Bennet's Agar (HI-MEDIA) M694

Yeast extract 1.000 HM peptone B # 1.000 Tryptone 2.000 Dextrose (GLucose) 10.000 Agar 15.000 Final pH (at 25°C) 7.3±0.2

ISP 7 medium (HI-MEDIA) M362

L-Asparagine 1.000 L-Tyrosine 0.500 Dipotassium hydrogen phosphate 0.500 Magnesium sulphate heptahydrate 0.500 Sodium chloride 0.500 *Trace salt solution (ml) 1.000 Agar 20.000:

*Trace salt solution contains - Ferrous sulphate heptahydrate 1.360mg Copper chloride, 2H2O 0.027mg Cobalt chloride, 6H2O 0.040mg Sodium molybdate, dihydrate 0.025mg Zinc chloride 0.020mg Boric acid 2.850mg Manganese chloride, tetrahydrate 1.800mg Sodium tartarate 1.770mg Final pH (at 25°C) 7.3±0.1

Appendix D
Software & Databases

Name	Executable	Description
CodonW	Windows	Program for codon and amino acid usage
XLSTAT	Windows	Statistical and data analysis software package
CMG-biotools	Linux	Stand alone OS for comparative microbial genomics
MEGA	Windows	Tool for sequence alignment and phylogeny
ClustalW	Windows	Multiple sequence alignment program
Bioedit	Windows	Biological sequence alignment editor
SPSS	Windows	Software package used for statistical analysis
FigTree	Windows	Graphical viewer of phylogenetic trees
BLAST	Windows/Linux	Algorithm for local similarity between sequences
R-Package	Windows/Linux	Functional language for statistical analysis
KyPlot	Windows	Software for plotting graphs

Name	Web Address	Description
JGI-IMG	www.img.jgi.doe.gov	Integrated Microbial Genomes system
NCBI	www.ncbi.nlm.nih.gov/	For molecular biology information
CAI Calculator2	http:// userpages.umbc.edu/ ~wug1/codon/cai/ cais.php	Calculation of codon adaptation index
RAST	rast.nmpdr.org	For the annotation of whole genome sequence
KEGG	www.genome.jp/ kegg	Kyoto Encyclopedia of genes and Genomes
DAMBE (v.5.0)	Windows	Software for sequence analysis
SwissDock	http:// www.swissdock.ch	To predict the molecular interactions that may occur between a target protein and a small molecule. ·
Open Babel	Windows/Linux	
MG RAST	http:// www.mg-rast.org	Phylogenetic and functional analysis of metagenomes

Appendix E

Abbreviations

%	Percent
/	Per
°C	Degree celsius
µmol l^{-1}	Micro mole per liter
3'→5'	3 prime to 5 prime
5'→3'	5 prime to 3 prime
v/v	Volume by volume
w/v	Weight by volume
α	Alpha
µg	Microgram
µl	Microlitre
µmol	Micromole
gm	Gram(s)
gm/l	Gram(s) per litre
hr	Hour(s)
ha	Hectare(s)
mg	Milligram
min	Minute(s)
ml	Millilitre(s)
mM	Millimolar
mm	Millimeter
µM	Micromolar
cm	Centimeter
kb	Kilo base pair
M	Molar

sec	Second(s)
rpm	Revolution per minute
ddH_2O	Double distilled water
O.D.	Optical Density
CLSI	Clinical Laboratory Standards Institute
NCCLS	National Committee for Clinical Laboratory Standards
CTAB	Cetyl Trimethyl Ammonium Bromide
DAS	Days after sowing
IS	Intermediately Susceptible
PGPR	Plant Growth Promoting Rhizobia
IAA	Indole-3-acetic acid
DNA	Deoxyribose Nucleic Acid
RNA	Ribose Nucleic Acid
PCR	Polymerase chain reaction
16SrRNA	16S ribosomal ribose nucleic acid
RAxML	(Randomnized Axelerated Maximum Likelihood, version
EMBL	European Molecular Biology Laboratory
NCBI	National Centre for Biotechnology Information
IMG	Integrated Microbial Genome
KEGG	Kyoto Encyclopedia of Genes and Genomes
CAI	Codon Adaptation Index
RSCU	Relative Synonymous Codon Usage
BLAST	Basic Local Alignment Search Tool
RAST	Rapid Annotations using Subsystems Technology

Index

National University Corporation
KAGOSHIMA UNIVERSITY 1-21-24 Korimoto, Kagoshima 890-0065 JAPAN http://www.kagoshima-u.ac.jp/
Since 1949

20th International Meeting on
Frankia and Actinorhizal Plants
29 – 31 May 2021
Online

Certificate of Participation

This is to certify that Dr. Gargi Sen participated in an online conference of the "20th International Meeting on *Frankia* and Actinorhizal Plants" organized by Ken-ichi Kucho (Kagoshima University), Takashi Yamanaka (Forestry and Forest Products Research Institute), Hiroyuki Tobita (Forestry and Forest Products Research Institute), Shunsuke Utsumi (Hokkaido University) and Toshiki Uchiumi (Kagoshima University) from 29 to 31 May 2021, and presented a paper on "Rhizospheric metagenomic analysis of *Alnus nepalensis* reveals the abundance of nitrogen-fixing symbiotic microbes" (oral presentation).

Ken-ichi Kucho
Graduate School of Science and Engineering,
Kagoshima University

Kenichi Kucho

Assam Botany Congress (ABC-02)
&
International Conference on Plant Science

(on blended mode)

3-5 December, 2021

Venue: Cachar College, Silchar, Assam

Organized by

Botanical Society of Assam, Guwahati

Department of Botany, Cachar College, Silchar, Assam

CERTIFICATE

Certified that **Ms. Gargi Sen** of **University of North Bengal** participated in the "Assam Botany Congress (ABC-02) and International Conference on Plant Science" held on 3-5 December, 2021 at Cachar College and delivered an oral presentation entitled **Microbial dynamics study depicts the importance of organic manure use in the tea garden soil.**

Dr. Tapan Dutta Secretary (BSA)	**Dr. Mukul Kr. Baruah** Organising Secretary	**Prof. Manabendra Dutta Choudhury** Convenor (Technical Program)	**Prof. Partha Pratim Baruah** Convenor (Technical Program)

JOURNAL OF FOREST RESEARCH
https://doi.org/10.1080/13416979.2022.2037813

ORIGINAL ARTICLE

Rhizospheric soil metabarcoding analysis of *Alnus nepalensis* from Darjeeling hills reveals the abundance of nitrogen-fixing symbiotic microbes

Gargi Sen[a], Indrani Sarkar[b], Saroja Chetrri[b], Pallab Kar[b], Ayan Roy[b], Arnab Sen[b,c] and Malay Bhattacharya[a]

[a]Department of Tea Science, University of North Bengal, Siliguri, India; [b]Bioinformatics Facility, Department of Botany, University of North Bengal, Siliguri, India; [c]Biswa Bangla Genome Centre, University of North Bengal, Siliguri, India

ABSTRACT

Actinorhizal plants are employed as successional plants for ecological restoration mainly due to their nitrogen-fixing ability. *Alnus nepalensis* or Nepali alder of Darjeeling hills is one of the potential plants used for agroforestry and known for its symbiotic association with an actinobacterium, *Frankia*. In this study, we performed a comparative16S rRNA amplicon analysis among six soil samples of *Alnus* rhizosphere and Non-rhizosphere of different altitudes of Darjeeling hills. Bioinformatics analyses were performed through the MG-RAST web server. Results revealed a set of 32 core bacterial genera among both rhizosphere and non-rhizosphere. Interestingly, *Alnus* rhizospheric soil samples were more populated with nitrogen-fixing taxa than non-rhizospheric or bulk soil. Nitrogen-fixing bacteria like *Frankia* and *Cyanobacteria* may play an important synergistic role in the proper growth and developmental stages of these plants. They are crucial for increasing the nitrogen amount of the soil through nitrogen fixation and gradually help in increasing the soil fertility and thus help in the proper progression of *Alnus* through the seral stages of succession. Nonrhizospheric soil samples were having a distinct population of other soil bacteria like *Streptomyces, Rubrobacter, and Xanthomonas* with higher Alpha diversity (54.14) than *Alnus* rhizospheric soil (42.18). This result was also validated by the rare fraction curve, which indicated more biodiversity in Non-rhizospheric soil rather than *Alnus* rhizosphere.

ARTICLE HISTORY
Received 6 July 2021
Accepted 30 January 2022

KEYWORDS
Metabarcoding; rhizosphere; alpha diversity; nitrogen fixation; *Alnus nepalensis*; *Frankia*

Introduction

Soil is one of the prime requisites for life on earth, which is a natural medium for the growth of plants, microbes, and other organisms. Soil can be considered as an ecosystem itself, as they support a variety of life and all accompanying complex interactions among organisms (Curtis and Sloan, 2005). The density of the microbial population in the soil is highly influenced by the interaction between host plants and microbes in the rhizosphere. The rhizodeposits (e.g. exudates, border cells, mucilage, etc.) are the major factors regulating the diversity and function of microbes on plant roots. The interactions are mainly mediated by Phyto signals released by the host plant (Mendes et al., 2018). The plants may moderate the microbiome of the rhizosphere to their advantage by selectively stimulating microorganisms with qualities that are beneficial to plant growth and health (Cook et al., 1995). A definite understanding of the microbiota of an ecosystem helps in sustainable agriculture, ecological restoration, reclamation of land, pathogen resistance, nutrient acquisition, and stress tolerance of the host plant (Liu et al., 2019). However, the traditional approach of microbiology fails to tap the microbial resources in their entirety and was not competent enough to provide clear estimates of the diversity. Hence, the study of microbial diversity using culture-independent approaches becomes necessary. In this context, one good approach is 16S metabarcoding analysis. It aids in the direct analysis of genomes contained within an environmental sample by providing sequence data of microbial communities as they exist in nature. It is a promising approach in describing the functional potential of the soil microbial community, which might yield greater insight into the health of soil than taxonomy-based metrics (Mardanov et al. 2018).

The present study focuses on 16S metabarcoding of soil samples from the *Alnus* rhizosphere and non-rhizosphere of natural forest areas of different altitudes of Darjeeling hills, India. The selected study regions of Darjeeling hills are part of the Siwalik Range or Outer Himalayas. The nature of the soil varies from loam to clayey loam and sandy loam. The main vegetation comprises tall and large trees like Alder (*Alnus*), Himalayan Birch, Alpine Fir, Oak, Chestnut, Walnut, Maple, etc. *Alnus* is an actinorhizal plant, a pioneer species, and acts as the primary successional plant in degraded habitats. *Alnus* belongs to the family Betulaceae which is locally known as "Utis", provides wood for poles, fuel, etc. Alders are fast-growing, show vigorous growth even in acidic soil and damaged sites such as burned areas and mining sites, further adding to their importance as the species of choice in forest restoration programs (Rana et al., 2018). *Alnus* plantation with cardamom (Sharma et al., 2002; Mukherjee, 2012), tea (Mortimer et al., 2015), and mandarin oranges as shade trees (Duke, 1983) signify its importance in the agroforestry system. The common Alder variety present in Darjeeling hills is *Alnus nepalensis,* which is widely distributed throughout the hills. Another species *Alnus nitida* is highly restricted to moist sandy areas and is rare (Shaw et al., 2014). *Alnus* is mainly associated with *Frankia, an* actinobacterium that is involved in nitrogen

CONTACT Gargi Sen sengargi@nbu.ac.in Department of Tea Science, University of North Bengal, Siliguri 734013, India
Supplemental data for this article can be accessed here.

fixation. The symbiont *Frankia* uses carbohydrates from alder trees to convert atmospheric N_2 into reactive nitrogen, a nitrate form directly available to plants (Huss-Danell et al., 1997), thereby enriching the soil.

Our current study is an attempt to assess the distribution and variation of microbes in the *Alnus* rhizosphere to the non-rhizosphere of different altitudinal regions of Darjeeling hills. This will be probably the first report of *Alnus nepalensis* rhizospheric soil 16S amplicon sequencing from India.

Material and methods

Sample collection

The sampling site was Darjeeling hills, West Bengal, India, with an altitudinal range that varies from 130 to 3660 m and covers an area of 1,200 sq. miles surrounded by the majestic peaks of Himalaya, characterized by a subtropical highland climate. The temperature varies from a minimum of −4°C to 25°C and the average annual precipitation of 3620 mm. Composite soil samples containing at least 3 sub-samples from the depth of 5–25 cm (collected in close proximity of around 1–2 meters) were taken from three locations belonging to different altitudes of Darjeeling hills from *Alnus* rhizosphere (R) and bulk soil (minimum distance of 10 meters from *Alnus*) or non-rhizospheric soil (NR) (Figure 1). Detailed sampling data including parameters like altitude, latitude, longitude, etc., and other biochemical factors of the soil are summarized in the Table 1. Sampling was performed in August 2019, a period in which the average temperature was about 18–21°C and the precipitation was 65 mm. The collection of soil samples was on a rainy day and nodules were not found. The samples were collected in sterile polythene bags and stored at 4°C till further processing.

DNA extraction

The soil samples (1 g each) were subjected to DNA extraction by using the MoBioPowerSoil® DNA Isolation Kit (MoBio Laboratories Inc., Carlsbad, CA, USA; cat # 12888–100) as per the manufacturer's protocol. The purified DNA in

Figure 1. Map of the study area.

Table 1. Organic matter and organic carbon were analysed after the procedures of Walkley and Black (1934); the amounts of nitrogen, phosphorus and potassium were determined after the methods of Jackson (2005).

Sample name	Place	Altitude(m)	Latitude	Longitude	Soil Temp.	pH	OC(%)	OM(%)	N(%)	P(ppm)	K(ppm)
R1	Ghoom	2282	27°0′11″ N	88°0′13″ E	18°C	3.62	3.28	5.64	0.28	16.1	214
NR1	Ghoom	2282	27°1′15″ N	88°0′9″ E	18°C	3.68	3.26	5.6	0.26	16	194
R2	Pashupatinagar	1877	26°0′55″ N	88°8′48″ E	20°C	3.71	2.35	4.04	0.21	15.5	156
NR2	Pashupatinagar	1877	26°1′25″ N	88°7′68″ E	20°C	3.73	2.26	3.88	0.18	15.3	145
R3	Mirik	1650	26°52′54″ N	88°11′19″ E	21°C	3.74	2.74	4.71	0.25	15.3	85.4
NR3	Mirik	1650	26°72′34″ N	88°15′29″ E	21°C	3.92	2.52	4.33	0.22	15.3	87.3

R: Rhizosphere, NR: Non-rhizosphere, OC: organic carbon, OM: Organic matter, N: Nitrogen, P: Phosphorous, K: Potassium

triplicate was pooled into a single sample to obtain enough DNA that collectively represented the microbial community composition in the soil samples. The quality and concentration of the pooled DNA from each sample were determined by using a Lambda II spectrophotometer (Perkin Elmer, Norwalk, Conn.) followed by agarose gel electrophoresis (1% w/v agarose in 1XTris-Acetate-EDTA (TAE) buffer, pH 7.8). Then, the purified DNA samples were dried and stored at 4°C until further use. Forty nanograms of extracted DNA was used for amplification along with 10 pM of each primer. The DNA was subjected to 16S amplicon sequencing targeting hypervariable V3-V4 region using 5' AGAGTTTGATGMTGGCTCAG 3' primer for forward sequence and 5′TTACCGCGGCMGCSGGCAC 3' primer for reversed sequence. Initial denaturation was done at 95°C followed by 25 cycles with denaturation at 95°C for 15 seconds, annealing at 60°C for 15 sec, and elongation at 72°C for 2 mins. Final extension at 72°C for 10 mins and hold at 4°C. 16S amplicon sequencing analysis has been performed using the Parallel-META pipeline (version 3.5; Jing et al., 2017).

16S rRNA sequencing and analysis pipeline

The 16S rRNA sequencing was done using the Illumina™ Nextseq platform (paired-end, 2 × 250 mode). The raw reads qualities were checked using FastQC. Sequence reads with >5 low-quality base pairs (<15 Phred score) were removed. Mean quality score for each base, per sequence quality score, per sequence GC contents and per base, N contents are calculated. Post QC, adapters are identified using the BBmerge function from the BBmap tool. Reads are trimmed with BBDuk tool, followed by quality check again. Post trimming, the quality check was done with the Fastqc tool for adapter content and reading the content of fastq files). Filtered reads were arranged into operational taxonomic units (OTUs) using Kraken v1.2.3 via The Galaxy Project (URL: usegalaxy.org) (Wood and Salzberg, 2014; Afgan et al., 2018). The phylogenetic architecture of all reference sequences is built by FastTree (Price et al., 2010). Since the 16S rRNA gene copy number varies greatly among different bacterial species, Parallel-META 3 (version 3.5; Jing et al., 2017) also calculates the precise relative abundance of each organism by 16S rRNA copy number calibration using IMG database (Markowitz et al., 2012). Besides, considering that the uneven sequencing depth (number of sequences) among multiple samples may introduce bias in detecting diversity patterns (Koren et al., 2013), an optional sequence rarefaction for sequencing depth normalization at the OTU level is provided after the taxonomic profiling. MG-RAST server (version 2.0) was used to analyze the high-quality reads from each sample for taxonomic profiling. The sequence data were uploaded to MG-RAST for further processing. For taxonomic annotation of sequence reads, the RefSeq database (Pruitt et al., 2005) was used and for functional gene annotation, and the SEED database (Overbeek et al., 2014) was used. Sequences were annotated using default settings of MG-RAST (maximum e-value cutoff of 10^{-5}, minimum % identity cut-off of 60%, and minimum alignment length cut-off of 15 bp). Subsystem Function level (the finest level) data was sub-sampled to 963693 bp reads per sample for diversity analysis. All the comparison was done following Spearman's Rank statistical analysis with $p < 0.01$.

Results

Sequence processing, quality filtering, and annotation

FastQC report revealed good quality reads indicating successful amplicon sequencing. Read adapters were trimmed and post trimming adapter content fell between 0% and 1% for all samples. Following quality check with FastQC tool, the paired-end reads from soil samples gave 54–56% average GC content and reads were 0.3–0.4 Million sequence reads with duplication values ranging from 47.30% to 82.90% (Table 2).

Taxonomic abundance analysis

Out of the six samples under study, the major phyla present in the rhizosphere (R1, R2, R3) and bulk soil or non-rhizosphere (NR1, NR2, NR3) are Proteobacteria, Bacteroidetes, Acidobacteria, Verrucomicrobia, Planctomycetes, Actinobacteria, Chloroflexi, Firmicutes, OD1, TM7, and others. Proteobacteria was the most abundant phylum (37.25%) with no significant difference in the distribution in the rhizosphere and non-rhizosphere region of different altitudes. This phylum is an important component of the soil microbiome, which is involved in biogeochemical cycles (Hunter et al., 2006). Bacteroidetes was the second abundant phylum with a distribution of 10–12% among all samples, but for the non-rhizosphere of Mirik (NR3), its population was exceptionally high (35%) ($p < 0.001$). It is a known fact that this phylum is adapted to different ecological niches and is a part of the gut microbiome (Moore and Holdeman, 1974; Leng et al., 2011). Interestingly, the increase in the Bacteroidetes population in NR3 area drastically reduced the presence of other phyla (Supplementary figure 1). It is also observed that the population of Acidobacteria, Verrucomicrobia, and Planctomycetes were reducing with a decrease in altitudes in the non-rhizosphere region, but it remained more or less the same in the rhizosphere region of different altitudes (Supplementary figure 1). The Actinobacteria phylum represented a uniform distribution in the rhizospheric region irrespective of the different altitudes (Supplementary figure 1).

At the genus level, we obtained a total of 2728 genera (apart from the unclassified taxa) from the six samples (R1, R2, R3, NR1, NR2, NR3). From Table 3, it is evident that the genera *Nocardia, Isosphaera, Gemmata, Gemmatimonas*, etc., including *Frankia* were the most abundant common genera with uniform distribution in the rhizosphere region of *Alnus* irrespective of their difference in altitudes. This

Table 2. Number of reads after adapter trimming and filtering for high-quality bases.

Sample Name	%age Duplicates	GC %	Mseqs
R1_Rd1	38.50%	55	0.1
R1_Rd2	89.30%	53	0.1
R2_Rd1	40.20%	55	0.1
R2_Rd2	89.50%	54	0.1
R3_Rd1	39.30%	54	0.1
R3_Rd2	89.50%	53	0.1
NR1_Rd1	39.50%	55	0.1
NR1_Rd2	88.80%	54	0.1
NR2_Rd1	39.70%	54	0.1
NR2_Rd2	90.00%	54	0.1
NR3_Rd1	41.10%	54	0.1
NR3_Rd2	90.50%	52	0.1

R: Rhizosphere, NR: Non-rhizosphere, Rd: Read

Table 3. Major microbial communities in the rhizosphere and non-rhizosphere.

R-1	Count	R-2	Count	R-3	Count	NR-1	Count	NR-2	Count	NR-3	Count
Nocardia	3413	***Nocardia***	3420	***Nocardia***	3416	*Xanthobacter*	2879	*Xanthomonas*	2222	*Xanthomonas*	2225
Isosphaera	2223	*Isosphaera*	2230	*Isosphaera*	2226	*Streptomyces*	1849	*Rubrobacter*	2122	*Rubrobacter*	2125
Gemmata	2133	*Gemmata*	2140	*Gemmata*	2136	*Rubrobacter*	1780	*Streptomyces*	1888	*Gemmata*	1848
Cyanobacteria	1780	***Cyanobacteria***	1787	***Cyanobacteria***	1783	*Polaromonas*	1580	*Ureaplasma*	1561	*Streptomyces*	1824
Gemmatimonas	1580	***Gemmatimonas***	1587	***Gemmatimonas***	1583	*Sideroxydans*	1561	*Thermodesulfo bacterium*	1511	*Borrelia*	1564
Clostridium	1561	*Clostridium*	1568	*Clostridium*	1564	*Thiobacillus*	1511	*Bilophila*	1457	***Cyanobacteria***	1514
Corynebacterium	1511	*Corynebacterium*	1518	*Corynebacterium*	1514	*Curvibacter*	1411	*Anaeroplasma*	1205	***Gemmatimonas***	1208
Burkholderia	1050	*Burkholderia*	1057	*Burkholderia*	1053	*Methylobacillus*	1050	*Kosmotoga*	1050	*Anaeroplasma*	1053
Rhodopirellula	798	*Rhodopirellula*	805	*Rhodopirellula*	801	*Candidatus* Accumulibacter	798	*Thermosipho*	879	*Clostridium*	882
Frankia	637	***Frankia***	644	***Frankia***	640	*Desulfatibacillum*	617	*Petrotoga*	798	*Ureaplasma*	801
Acidobacterium	631	*Acidobacterium*	638	*Acidobacterium*	634	*Bilophila*	614	*Rubritalea*	621	*Corynebacterium*	624
Staphylococcus	614	*Staphylococcus*	621	*Staphylococcus*	617	*Desulfotalea*	611	*Aspergillus*	526	*Thermodesulfobact.*	617
Saccharopolyspora	536	*Saccharopolyspora*	543	*Candidatus* Cloacamonas	539	*Desulfuromonas*	516	*Paracoccidioides*	460	*Burkholderia*	529
Pseudomonas	460	*Pseudomonas*	467	*Vaccinium*	463	*Pelobacter*	460	*Nectria*	426	*Kosmotoga*	463
Ktedonobacter	426	*Ktedonobacter*	433	*Bacteroides*	429	*Corallococcus*	426	*Trachelomonas*	282	*Rhodopirellula*	429

R: Rhizosphere, NR: Non-rhizosphere; (Diazotroph taxa are in bold)

indicates that a specific microbial community is being attracted towards the rhizospheric region and also found that the majority of the core microbial community in the soil shows nitrogen-fixing properties. On contrary, the non-rhizosphere region exhibited no specific choice in microbial population and had a high diversity index (Supplementary Figures 2–4). We have also measured the soil temperature at different collection sites. However, taxon distribution was not found to be influenced by the temperature of the soil.

Rhizospheric effects on microbiome diversity

Based on the OTU number, Chao1 bacterial species abundance index, and Shannon microbial diversity index, α-diversity analysis was conducted on various samples. Results indicate that indigenous microbial community diversity was significantly higher in non-rhizospheric soil than *Alnus* rhizosphere, which holds good for all three altitudes. This validates the previous report that shows rhizospheric diversity is lower than bulk soil (Costa et al., 2006; Hein et al., 2008). A significant rhizosphere effect was reflected by reduced microbiome diversity in the *Alnus* rhizosphere compared with that of the non-rhizosphere. The Simpson index of alpha diversity for the rhizosphere ranges was found to be 76.31–125, and in the non-rhizosphere, it varies from 25.38 to 69.45. The microbial diversity of the rhizosphere decreases from high to low altitudes, but intriguingly, diversity in non-rhizosphere regions decreases from low to high altitudes.

The assessment of the Beta diversity of microbial communities across different rhizosphere and non-rhizospheric soil indicates the effect of altitudes among the microbial population of the rhizosphere and non-rhizosphere. On the functional front, both the altitudes and soil conditions such as temperature and pH play an equal role in microbial diversity. PCA analysis was used to study the similarity among various rhizosphere and bulk soil samples in the structures of bacterial communities exploring the main influencing factors driving the differences in micro-community compositions. The axis in the PCA plot indicated the total sequence reads for each soil sample. The subsequent multivariate analysis shows the clustering of non-rhizosphere and rhizosphere on the same axis, indicating similar environmental conditions of the soil. Nonetheless, the formation of separate clusters in the same axis demonstrates the specific choice of the microbial community in these regions (Figure 2). The rarefaction curve for the rhizosphere and non-rhizosphere (Figure 3) shows that the microbial population of rhizosphere soil is getting saturated at a higher level (5000 sequence reads) compared to the non-rhizosphere (3500 sequence reads). This is indicative of the volatile microbiome of the *Alnus* rhizosphere due to the host plant's interaction with the microbial population. Fascinatingly, it is observed that the non-rhizosphere presents a stable scenario where it gets saturated at a lower level (Figures 2 and 3).

The functional profiling analysis showed that KEGG Ontology (KOs) were mainly involved in 23 KEGG level 2 pathways and 25 KEGG level 3 pathways for both rhizosphere and non-rhizosphere. This may be because microbes are evolved in the same physical environment, which has resulted in the lack of variation in metabolic pathways.

The microbial interaction network is constructed to explore co-occurrence and co-exclusion patterns of organisms or functional genes across microbial communities. In the interaction network, each node represents a single organism (or gene), and nodes are connected by links that represent their correlation coefficient of abundance variation among multiple samples (Faust and Raes, 2012). The network analysis shows that the major phyla were represented as 23 nodes belonging to Proteobacteria, Cyanobacteria, Nitrospirae, Firmicutes, Acidobacteria, Planctomycetes, Actinobacteria, Verrucomicrobia, Chloroflexi, Bacteroidetes, etc. Furthermore, when we analyzed the bacterial count of rhizosphere versus non-rhizosphere (bulk soil), we found that approximately 50% of the microbial population was associated with nitrogen metabolism

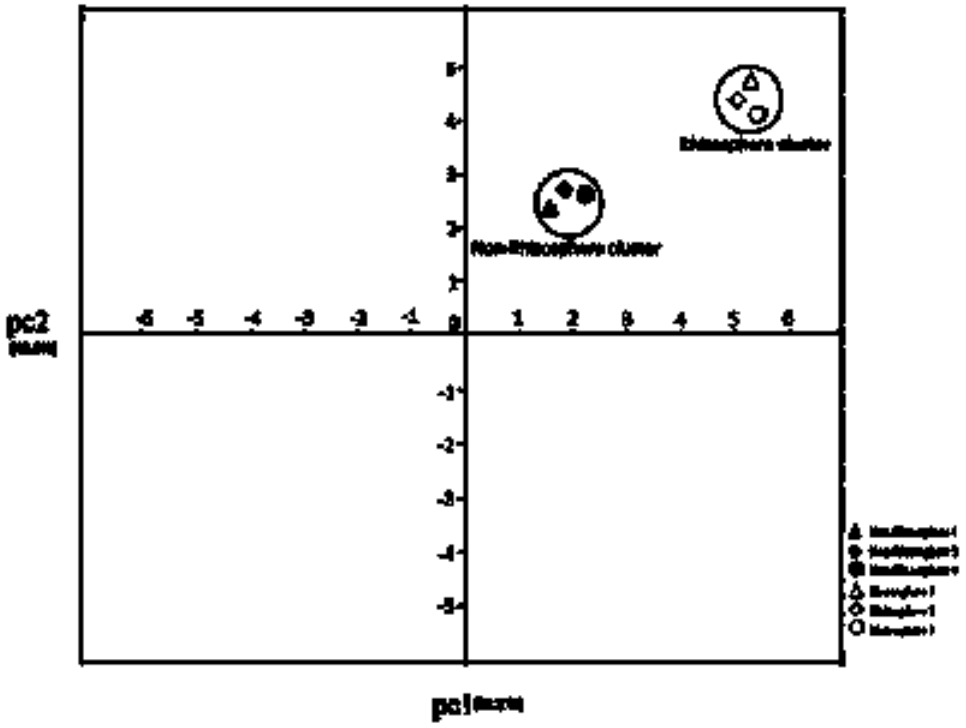

Figure 2. PCA plot for rhizospheric and non rhizospheric sample.

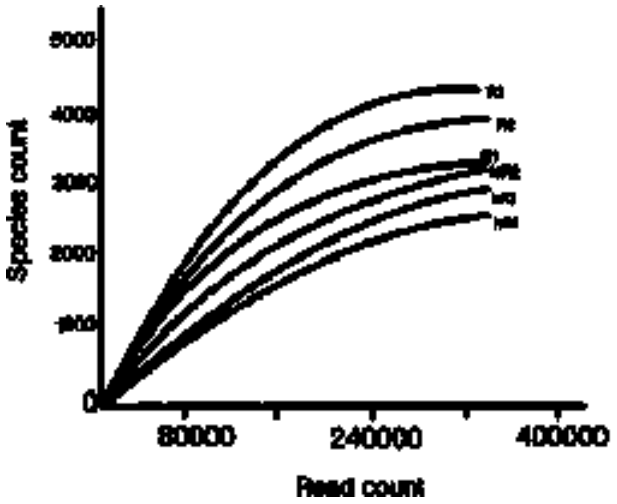

Figure 3. Rarefaction curve for rhizospheric and non rhizospheric sample.

in the rhizosphere (Table 3). This substantiates our findings that the *Alnus* rhizosphere is enriched with a specific microbial community that exhibits nitrogen fixation properties. Contrary to that, the non-rhizospheric soil exhibit rather neutral behavior in the case of microbial populations. This corroborates with the earlier findings where allelochemical inhibition of nitrification in nondiazotrophic plants was reported by Rice and Pancholy (1972) and Gökceoğlu (1988).

Discussion

In this present study, we compared the microbial diversity of *Alnus* rhizospheric and non-rhizospheric soil from different altitudes of Darjeeling hills, India, using amplicon sequencing approaches. Earlier reports show that though there is great diversity in the bacterial communities, very few taxa are predominant in any given soil sample (Xu et al., 2018). In consensus with this, we have also found a few bacterial taxa, such as Proteobacteria, Bacteroidetes, Acidobacteria, Actinobacteria, Planctomycetes, Verrucomicrobia, etc., that are abundant in both the bulk soil and *Alnus rhizosphere*. This trend of selective abundance was also found by other workers (Mendes et al., 2014; Edwards et al., 2015). Bacteroidetes, Gemma timonadetes, and all proteobacterial classes *(α, β, δ, γ)* were identified as potential copiotrophs, which are involved in carbon metabolism (Elliott et al. 2014). Further analysis showed that the diversity index of the rhizosphere is more in higher altitudes than the lower altitudinal levels. However, the reverse scenario was observed in the non-rhizospheric region. One possible explanation could be that human intervention is reduced at higher altitudes, so the specific microbiome nurtured by the actinorhizal plant, *Alnus* forms a stable community. This is in agreement with the earlier studies that, root exudates from plant species select specific microbial populations in the rhizosphere region (Carvalhais et al., 2011). As the altitude decreases, there is more intrusion from man and animals, which disturbs this equilibrium. The reverse condition in the non-rhizosphere could be because of the lack of host-specific interactions and the change in the available organic matter due to the wide range of involvement of domestic animals, humans, etc. It has already been mentioned that the Bacteroidetes were found to be abundant in the non-rhizosphere soil of Mirik (NR3), which is involved in cellulose degradation and is a common gut microbe. Soil pH increased from high to low altitudes, which also correlates with the reducing population of *Acidobacteria*, an oligotroph in lower altitudes. Different bacterial taxa such as *Proteobacteria, Bacteroidetes, Firmicutes, Actinobacteria*, etc., coexist in the soil indicating their minimal competition for resources. The abundance of nitrogen-fixing and nitrogen metabolizing microbes in the *Alnus* rhizosphere underlines the fact that the host plant is instrumental in maintaining distinct rhizobiome. This corroborates with the previous reports that a plant's root exudates can influence the diversity of microorganisms (Igiehon and Babalola, 2017, 2018). Another interesting observation was that *Frankia* was found to be associated with several nitrogen-fixing microbes in the rhizosphere. This may be an indication of their evolutionary association, where the nitrogen-fixing microbes might have existed as symbiotic organisms in the past. As the evolution progressed, they could have established themselves as free-living but tend to remain in the *Alnus* rhizosphere. This association of *Frankia* and other microbes as well as the host-induced signals may act synergistically to create a specific microbiome for the *Alnus* rhizosphere.

Conclusion

In the present study, we intended to assess the microbial populations of the *Alnus* rhizosphere and that of non-rhizosphere (bulk soil) in different altitudes of Darjeeling Hills. We performed meta-barcoding of the soil samples and found that despite altitudinal variations, the *Alnus* rhizosphere was found to be enriched with the diazotrophs, whereas the bulk soil was relatively neutral in this regard. This study is probably the maiden comparative analysis of the *Alnus nepalensis* rhizosphere and non-rhizosphere of varying altitudes, which is purely based on the soil metabarcoding data. However, this is a very interesting topic and demands detailed study on the aspects of nodule formation, diversity of taxa in nodule and rhizosphere, etc., which was beyond the scope of the present study.

Acknowledgments

We acknowledge Mr. Moinak, Soil testing laboratory, Department of Tea Science, NBU and University of North Bengal (NBU) for supporting this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Arnab Sen http://orcid.org/0000-0002-2079-5117
Malay Bhattacharya http://orcid.org/0000-0002-4066-2979

Authors' contribution

GS, conceived the idea. GS, SC, PK, AS, MB did the soil collection and preparation. GS, IS & AR performed the Bioinformatic analysis. All authors contributed to the manuscript preparation and approved the final draft.

References

Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, et al. 2018. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46(W1):W537–W544. doi:10.1093/nar/gky379.

Carvalhais LC, Dennis PG, Fedoseyenko D, Hajirezaei MR, Borriss R, von Wirén N. 2011. Root exudation of sugars, amino acids, and organic acids by maize as affected by nitrogen, phosphorus, potassium, and iron deficiency. J Plant Nutr Soil Sci. 174(1):3–11. doi:10.1002/jpln.201000085.

Cook RJ, Thomashow LS, Weller DM, Fujimoto D, Mazzola M, Bangera G, Kim DS. 1995. Molecular mechanisms of defense by rhizobacteria against root disease. Proc Natl Acad Sci USA. 92 (10):4197. doi:10.1073/pnas.92.10.4197.

Costa R, Gotz M, Mrotzek N, Lottmann J, Berg G, Smalla K. 2006. Effects of site and plant 663 species on rhizosphere community structure as revealed by molecular analysis of microbial 664 guilds. FEMS Microbiol Ecol. 56(2):236–249. doi:10.1111/j.1574-6941.2005.00026.x.

Curtis TP, Sloan WT. 2005. Exploring microbial diversity–A vast below. Science. 309(5739):1331–1333. doi:10.1126/science.1118176.

Duke JA. 1983. Handbook of energy crops. West Lafayette (IN): Purdue University Press.

Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. Proc Natl Acad Sci USA. 112(8):E911–E920.

Elliott DR, Thomas AD, Hoon SR, Sen R. 2014. Niche partitioning of bacterial communities in biological crusts and soils under grasses, shrubs and trees in the Kalahari. Biodivers Conserv. 23 (7):1709–1733. doi:10.1007/s10531-014-0684-8.

Faust K, Raes J. 2012. Microbial interactions: from networks to models. Nat Rev Microbiol. 10(8):538–550. doi:10.1038/nrmicro2832.

Gökceoğlu M. 1988. Nitrogen mineralization in volcanic soil under grassland, shrub and forest vegetation in the Aegean region of Turkey. Oecologia. 77(2):242–249. doi:10.1007/BF00379193.

Hein JW, Wolfe GV, Blee KA. 2008. Comparison of rhizosphere bacterial communities in 670 Arabidopsis thaliana mutants for systemic acquired resistance. Microb Ecol. 55:333–343. doi:10.1007/s00248-007-9279-1.

Hunter EM, Mills HJ, Kostka JE. 2006. Microbial community diversity associated with carbon and nitrogen cycling in permeable shelf sediments. Appl Environ Microbiol. 72(9):5689–5701. doi:10.1128/AEM.03007-05.

Huss-Danell K, Uliassi D, Renberg I. 1997. River and lake sediments as sources of infective Frankia (Alnus). Plant Soil. 197(1):35–39. doi:10.1023/A:1004268931699.

Igiehon NO, Babalola OO. 2017. Biofertilizers and sustainable agriculture: exploring arbuscular mycorrhizal fungi. Appl Microbiol Biotechnol. 101(12):4871–4881. doi:10.1007/s00253-017-8344-z.

Igiehon NO, Babalola OO. 2018. Rhizosphere microbiome modulators: contributions of nitrogen fixing bacteria towards sustainable agriculture. Int J Environ Res Pub Health. 15(4):574. doi:10.3390/ijerph15040574.

Jackson ML. 2005. Soil chemical analysis: advanced course. Madison (WI): UW-Madison Libraries, Parallel Press.

Jing G, Sun Z, Wang H, Gong Y, Huang S, Ning K, Xu J, Su X. 2017. Parallel-META 3: comprehensive taxonomical and functional analysis platform for efficient comparison of microbial communities. Sci Rep. 7:40371. doi:10.1038/srep40371.

Koren O, Knights D, Gonzalez A, Waldron L, Segata N, Knight R, Huttenhower C, Ley RE. 2013. A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. PLoS Comput. Biol. 9(1): e1002863.

Leng J, Xie L, Zhu R, Yang S, Gou X, Li S, Mao H. 2011. Dominant bacterial communities in the rumen of Gayals (*Bos frontalis*), Yaks (*Bos grunniens*) and Yunnan Yellow Cattle (*Bos taurus*) revealed by denaturing gradient gel electrophoresis. Mol Biol Rep. 38 (8):4863–4872. doi:10.1007/s11033-010-0627-8.

Liu F, Hewezi T, Lebeis SL, Pantalone V, Grewal PS, Staton ME. 2019. Soil indigenous microbiome and plant genotypes cooperatively modify soybean rhizosphere microbiome assembly. BMC Microbiology. 19(1):201. doi:10.1186/s12866-019-1572-x.

Mardanov AV, Kadnikov VV, Ravin NV. 2018. Metagenomics: a Paradigm Shift in Microbiology. In: Nagarajan M, editor. Metagenomics: perspectives, Methods, and Applications. Cambridge (MA): Academic Press; p. 1–13.

Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M. 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40(D1):D115–D122.

Mendes LW, Kuramae EE, Navarrete AA, Van Veen JA, Tsai SM. 2014. Taxonomical and functional microbial community selection in soybean rhizosphere. ISME J. 8(8):1577–1587. doi:10.1038/ismej.2014.17.

Mendes LW, Raaijmakers JM, de Hollander M, Mendes R, Tsai SM. 2018. Comparison of rhizosphere bacterial communities in Arabidopsis thaliana mutants for systemic acquired resistance. ISME J. 12(1):212–224. doi:10.1038/ismej.2017.158.

Moore WE, Holdeman LV. 1974. Special problems associated with the isolation and identification of intestinal bacteria in fecal flora studies. Am J Clin Nutr. 27(12):1450–1455. doi:10.1093/ajcn/27.12.1450.

Mortimer PE, Gui H, Xu J, Zhang C, Barrios E, Hyde KD. 2015. Alder trees enhance crop productivity and soil microbial biomass in tea plantations. Appl Soil Ecol. 96:25–32.

Mukherjee D. 2012. Resource conservation through indigenous farming system in hills of West Bengal. J Crop Weed. 8(1):160–164.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, et al. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 42(D1):D206–D214.

Price MN, Dehal PS, and Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One. 5(3):e9490.

Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 33(suppl–1):D501–D504.

Rana SK, Rana HK, Shrestha KK, Sujakhu S, Ranjitkar S. 2018. Determining bioclimatic space of Himalayan alder for agroforestry systems in Nepal. Plant Divers. 40(1):1–18.

Rice EL, Pancholy SK. 1972. Inhibition of nitrification by climax ecosystems. Am J Bot. 59(10):1033–1040.

Sharma G, Sharma R, Sharma E, Singh KK. 2002. Performance of an age series of Alnus–cardamom plantations in the Sikkim Himalaya: nutrient dynamics. Ann Bot-London. 89:273–282.

Shaw K, Stritch L, Rivers M, Roy S, Wilson B, Govaerts R. 2014. Betulaceae. Richmond (UK): Botanic Gardens Conservation International.

Walkley A, Black IA. 1934. An examination of the Degtjareff method for determining soil organic matter, & a proposed modification of the chromic acid titration method. Soil Sci. 37:29–38.

Wood DE, Salzberg SL. 2014. Kraken: ultrafast amplicon sequencing sequence classification using exact alignments. Genome Biol. 15:R46.

Xu J, Zhang Y, Zhang P, Trivedi P, Riera N, Wang Y, Liu X, Fan G, Tang J, Coletta-Filho HD, et al. 2018. The structure and function of the global citrus rhizosphere microbiome. Nat Commun. 9:4894.

In Silico Biology 7 (2007) 547–558
IOS Press

Analysis of Codon Usage Patterns and Predicted Highly Expressed Genes for Six Phytopathogenic *Xanthomonas* Genomes Shows a High Degree of Conservation

Gargi Sen[a], Saubashya Sur[a], Debadin Bose[a], Uttam Mondal[b], Teal Furnholm[c], Asim Bothra[b], Louis Tisa[c] and Arnab Sen[a,c,*]

[a]*DBT-Bioinformatics facility, Molecular Genetics Laboratory, Department of Botany, University of North Bengal, Siliguri 734013, India*
[b]*Bioinformatics Cheminformatics Laboratory, Department of Chemistry, Raiganj University College, Raiganj, India*
[c]*Department of Microbiology, University of New Hampshire, Durham NH, USA*

Edited by H. Michael; received 20 April 2007; revised 8 August 2007; accepted 9 August 2007; published 21 August 2007

ABSTRACT: Members of the genus *Xanthomonas* are significant phytopathogens, which cause diseases in several economically important crops including rice, canola, tomato, citrus, etc. We have analyzed the genomes of six recently sequenced *Xanthomonas* strains for their synonymous codon usage patterns for all of protein coding genes and specific genes associated with pathogenesis, and determined the predicted highly expressed (PHX) genes by the use of the codon adaptation index (CAI). Our results show considerable heterogeneity among the genes of these moderately G+C rich genomes. Most of the genes were moderate to highly biased in their codon usage. However, unlike ribosomal protein genes, which were governed by translational selection, those genes associated with pathogenesis (GAP) were affected by mutational pressure and were predicted to have moderate to low expression levels. Only two out of 339 GAP genes were in the PHX category. PHX genes present in clusters of orthologous groups of proteins (COGs) were identified. Genes in the plasmids present in two strains showed moderate to low expression level and only a couple of genes featured in the PHX list. Common genes present in the top-20 PHX gene-list were identified and their possible functions are discussed. Correspondence analysis showed that genes are highly confined to a core in the plot.

KEYWORDS: Codon usage, codon adaptation index, correspondence analysis, pathogenicity, *Xanthomonas*

INTRODUCTION

Members of the γ-proteobacteria genus *Xanthomonas* are plant-associated bacteria and most species are significant plant pathogens that cause disease in various economically important plants [1]. Among them, *X. oryzae* pv. *oryzae* is a pathogen of the staple crop plant rice (*Oryza sativa*), *X. axonopodis* pv. *citri* causes citrus bacterial canker in many citrus producing tropical and subtropical countries around the world [2], *X. campestris* pv. *vesicatoria* is responsible for bacterial spot or black spot of tomato or

*Corresponding author. E-mail: senarnab_nbu@hotmail.com.

Electronic publication can be found in *In Silico Biol.* **7**, 0039 <http://www.bioinfo.de/isb/2007/07/0039/>, 21 August 2007.

capsicum, while bacterial black rot of canola is caused by *X. campestris* pv. *campestris*. The genomes for all the above plant pathogens have been sequenced [3–6] and shown to contain circular chromosomes that have coding densities very similar to other sequenced bacteria. The availability of these complete genome sequences opens the door to the potential use of bioinformatics approaches that focus on the codon usage profile and to investigate gene expression and regulation in the context of global cellular network.

Codon usage and codon preferences vary significantly within and between organisms [7–9]. Across the genome, the G+C composition resulting from mutational bias and/or translational selection has been hypothesized to determine the major trends in codon usage by high or low G+C organisms [10]. Within a given genome, codon bias is much higher in highly expressed genes than in lowly expressed ones [11–15]. The bias of highly expressed genes is more affected by translational selection than the lowly expressed genes, which are directed by mutational bias [16]. To dissect the patterns and causality of codon usage, many indices have been proposed to measure the degree and direction of codon bias [12]. Among these indices, the codon adaptation index (CAI) was proposed as a measure of codon usage within a gene relative to a reference set of genes (usually ribosomal protein genes) [12]. This index has been shown to correlate best with mRNA expression levels [13]. Besides CAI, the effective number of codons (Nc), which is defined as the number of equal codons that would generate the same codon usage bias as observed, and the frequency of optimal codons (Fop), which is defined as the fraction of synonymous codons that are optimal codons, are also used for the same purpose. The codon bias index (CBI), which measures the extent that a gene uses a subset of optimal codons, is used as an indicator of the cause of codon bias. A low CBI value indicates that the biasness may be to mutational selection whereas elevated CBI may point to the translational efficiency as the cause [17,18].

Xanthomonas species are plant pathogens. Lee *et al*. [3] recognized three major groups of genes related to pathogenesis: (1) effector or avirulence genes (*avr*), (2) hypersensitive response and pathogenicity (*hrp*) genes, (3) the *gum* gene cluster. With phytopathogenic bacteria, a type III protein secretion system (TTSS) encoded by *hrp* genes plays a central role in eliciting defense responses, including rapid cell death response. On non-host or resistant host plants, this response is called the hypersensitive reaction (HR), and leads to pathogenesis on susceptible hosts [19]. Some Hrp proteins form a pilus that has been proposed to function as conduit to directly translocate effector proteins such as avirulance factor into plants [20]. In addition to the TTSS, the type II secretion system may play a role in secretion of other virulence factor with many *Xanthomonas* species, the *gum* gene cluster involved in exopolysaccharide synthesis functions as a virulence determinant [21]. Among the six sequenced *Xanthomonas* genomes, the *hrp* genes occur in frequencies of 2 to 20 while the *gum* genes and *avr* genes occur in frequencies of 13 to 16 and 3 to 18 respectively [3,4,19–21].

The aim of the present study was to perform a comparative analysis of the codon usage patterns and predicted expression levels for the protein coding genes in these phytopathogenic bacteria with special reference to those genes associated with pathogenesis.

METHODS

Genome sequences for six *Xanthomonas* strains (*Xanthomonas axonopodis* pv. *citri* 306, *Xanthomonas campestris* pv. *campestris* 8004, *Xanthomonas campestris* pv. *campestris* ATCC 3391, *Xanthomonas campestris* pv. *vesicatoria* 85-10, *Xanthomonas oryzae* pv. *oryzae* KACC10331and *Xanthomonas oryzae* pv. *oryzae* MAFF 311018 (hence forth, these strains will be referred to as XAC, XCC1, XCC2, XCV, XOO1 and XOO2, respectively) were obtained from the IMG website (http://img.jgi.doe.gov). Table 1

Table 1
Salient features of *Xanthomonas* genomes analyzed in this study

Organism	XAC	XCC1	XCC2	XCV	XOO1	XOO2
DNA, total number of bases	5274174	5148708	5076188	5420152	4941439	4940217
G+C content (%)	64.71	64.96	65.07	64.56	63.69	63.70
Genes total number	4487	4333	4240	4786	4140	4431
Protein coding genes	4427	4273	4181	4726	4080	4372
RNA genes	60	60	59	60	60	59
rRNA genes	6	6	6	4	6	6
tRNA genes	54	54	53	56	54	53
Genes with function prediction	2751	2664	2690	3389	2872	2776
Pseudogenes	0	0	0	0	0	0
Genes assigned to enzymes	593	583	587	905	486	506
Genes in COGs	3210	3133	3133	3286	2921	3059
Number of plasmids	2	0	0	4	0	0
Total number of genes in plasmids	115	0	0	241	0	0

XAC, *Xanthomonas axonopodis* pv. *citri* 306; XCC1, *Xanthomonas campestris* pv. *campestris* 8004; XCC2, *Xanthomonas campestris* pv. *campestris* ATCC 3391; XCV, *Xanthomonas campestris* pv. *vesicatoria* 85-10; X001, *Xanthomonas oryzae* pv. *oryzae* KACC10331; X002, *Xanthomonas oryzae* pv. *oryzae* MAFF 311018.

shows some of the general features of these genomes. Only two of the strains contain plasmids. Strain XAC has two plasmids, pXAC64 (64.9 kb) and pXAC33 (33.7 kb), which have 73 and 42 genes, respectively. Strain XCV maintains four plasmids, pXCV183 (182.6 kb), pXCV38 (38.1 kb), pXCV19 (19.1 kb), and pXCV2 (1.85 kb), which have 173, 43, 23 and 2 genes, respectively.

All of the protein coding genes and those genes associated with the ribosomal proteins, plasmids, virulence/avirulence related traits, hypersensitive response and pathogenesis (*hrp*), and *gum* gene clusters were identified from the available literature [3,4,19–21] and were analyzed by the use of CodonW software (http://sourceforge.net/projects/codonw/) and CAI Calculator 2 (http://www.evolvingcode.net/codon/CalculateCAIs.php).

The software CodonW [18,22] was used to calculate GC3s, Nc [23], RSCU [12], CBI, and Fop values [21,22]. GC3s symbolize the frequency of guanine and cytosine at the synonymous third positions of codons. The effective number of codons (Nc) is a simple measure of overall codon bias [23]. Its value represents the number of equal codons that would generate the same codon usage bias that was observed. Nc values range from 20 (when only one codon is per amino acid) to 61 (when all codons are used in equal probability). The relative synonymous codon usage (RSCU) is defined as the ratio of the observed frequency of a codon to the expected frequency if all the synonymous codons for those amino acids are used equally [12]. The values of RSCU, which are greater than 1, reveal that the corresponding codons are used more often than the expected frequency and *vice versa* [15]. The codon bias index (CBI) [24] is a quantum of directional codon bias and measures the extent to which a gene uses a subset of optimal codons. In a gene with extreme codon bias the CBI value may be equal to 1. Fop is the fraction of synonymous codons that are optimal codons. Its value ranges from 0 (meaning a gene has no optimal codons) to 1.0 (when a gene is entirely composed of optimal codons).

Another very widely used measure of codon bias in prokaryotes and eukaryotes is the codon adaptation index (CAI). It is a measurement of relative adaptiveness of a gene's codon usage towards the codon usage of highly expressed genes. The relative adaptiveness of each codon is the ratio of the usage of each codon, to that of the most abundant codon within the same synonymous family. The CAI value vary from 0 to 1.0 with higher CAI values indicating that the gene of interest has a codon usage pattern more similar to that in the reference genes.

Correspondence analysis (COA) was also performed using Codon W 1.4.2. This method investigates the major trends in codon and amino acid variations among the genes.

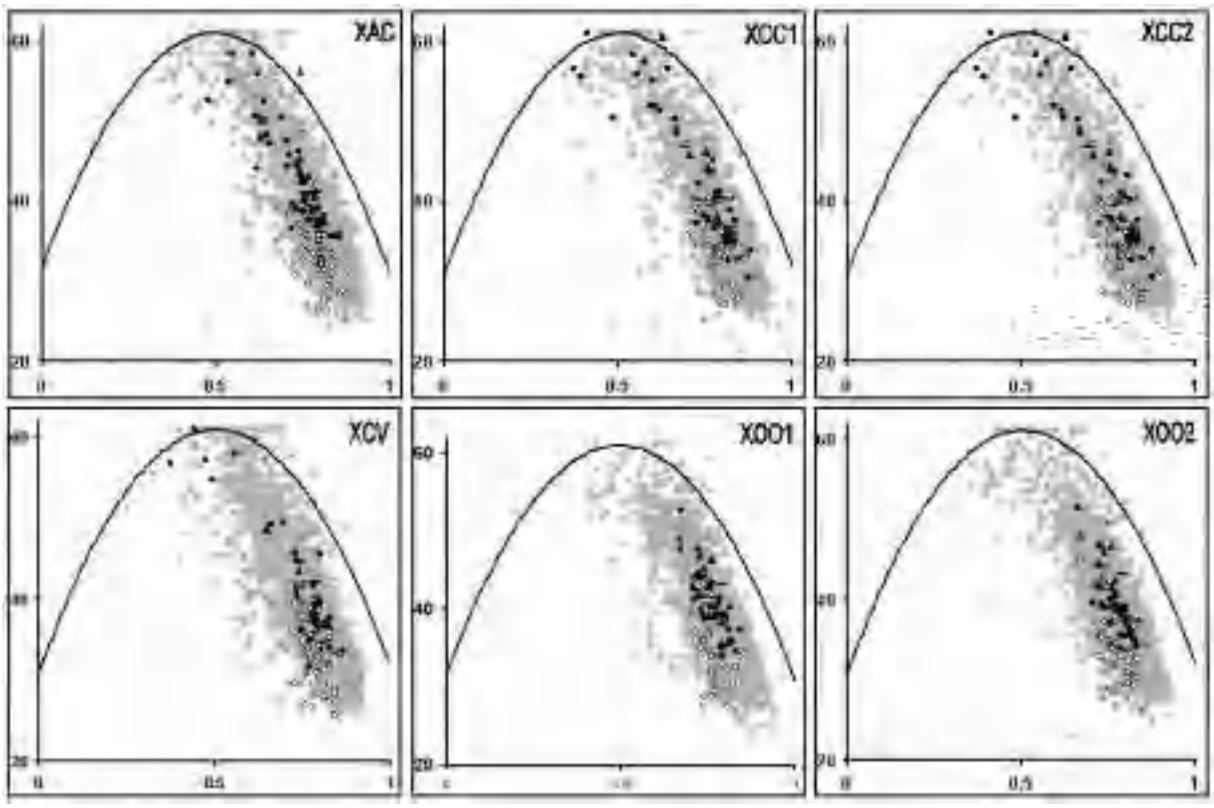

Fig. 1. Effective number of codons used (Nc) in each gene plotted against the G+C content at synonymous third position of codons (GC3) for all six *Xanthomonas* genomes. The continuous curve in each plot represents the null hypothesis that the GC bias at the synonymous site is solely due to mutation but not selection [23]. Gray circle = protein coding genes; hollow circle = ribosomal protein genes; '−' = *gum* genes; '▲ = *hrp* genes and '•' = *vir* genes. In all the plots x and y axis correspond to GC3 and Nc respectively.

RESULTS AND DISCUSSION

Heterogeneity in codon usage in Xanthomonas genomes

Our first aim in this study on the codon usage patterns among various *Xanthomonas* genomes was to determine the degree of heterogeneity in codon use. Most bacteria with a balanced AT/GC genome content have a considerable amount of codon heterogeneity [25]. Codon heterogeneity is usually associated with gene expression level. Thus, highly expressed genes contain a higher frequency of codons that are considered translationally optimal [13,14,25]. Since *Xanthomonas* have a moderately high G+C content, the GC3s and Nc values for all of the genes in these genomes were calculated to determine if codon heterogeneity exists among genes of various *Xanthomonas* species. The results from this analysis were plotted and shown in Fig. 1.

These Nc *vs* GC3s plots have been suggested to be an effective means to investigate the codon usage variations among genes in the same genome [22]. The Nc values of the *Xanthomonas* genes range from 22 ± 2 to 61 ± 0 for all genomes suggesting that these moderately GC-rich genomes exhibited considerable heterogeneity in codon usage. The genes encoding ribosomal proteins, which are expected to be expressed at high levels during rapid cell growth, were identified and are highlighted in the Nc

plots. Most of the ribosomal protein genes of all *Xanthomonas* genomes clustered at the low ends of the plot, which is similar to the results observed with genomes from *Escherichia coli* and two *Streptomyces* species [26], and indicate a significant strong codon bias in the ribosomal protein genes which was the result of selection for translational efficiency [27]. The location of the *gum*, *hrp*, and *vir* genes are also indicated in the Nc plots (Fig. 1). For the XAC, XCC1, XCC2 and XCV genomes, the *gum* and *hrp* genes were relatively clustered together with the ribosomal protein genes, except for one gene each in the XAC, XCCI and XCC2 genomes (in all three cases it was *HrpE* gene), while as expected considerable codon heterogeneity was found among the *vir* genes [26]. However, contrary to other four genomes, *vir* genes in XOO1 and XOO2 were clustered together. If GC composition is not the only factor influencing codon usage bias, analysis of the distribution of the genes in an Nc/GC3s plot would indicate the other factors. If synonymous codon bias was absolutely dictated by GC3s, Nc values should fall on the expected curve the GC3 and Nc plot [22]. However, we found that except for a very few genes, the values obtained for the majority of the genes were well below the expected values (Fig. 1). This result indicates that codon usage bias for the majority of *Xanthomonas* genes is affected independently of overall base composition.

Table 2 shows the mean values of different indices used to study codon usage patterns. Variation in the mean Nc values for the different gene groups was observed within the same species as well as other species. As expected, a correlation between Nc and GC3 was observed: Nc values decreased with a corresponding increase in GC3 values and *vice-versa*. These low Nc values indicated a high degree of codon bias. Ribosomal protein genes had a lower mean Nc value than the mean value obtained for all of the protein encoding genes for all of the genomes, the mean Nc values for the virulence and hrp genes were higher, except for the virulence genes of X001 and X002. Thus, ribosomal protein genes were more highly bias compared to those associated with pathogenesis.

An analysis of the Fop values for the different gene sequences among different species of *Xanthomonas* showed some variation (Table 2) The mean Fop values of the potentially highly expressed ribosomal protein genes were higher than the mean Fop of the total protein coding gene sequences for these genomes, while the mean Fop values of the different gene groups were lower. This result indicates that ribosomal protein genes have higher proportion of optimal codons. If mutational bias only influenced codon bias, these genes would have had a low Fop value. Since that was not the case for these *Xanthomonas* genomes, there may be other factors acting on codon bias. These data were analyzed by the use of the codon bias index (CBI) [24] which is a measure of directional codon bias and measures the extent to which a gene uses a subset of optimal codons. In our study, we found that the CBI values for the gene sequences were low in the range of 0.192 to 0.354 (Table 2). These low CBI values indicate lower level of biasness. Ribosomal protein genes had comparatively higher CBI values compared to other gene groups. CBI values correlated with Nc values similar to their correlation with GC3 values. Genes having lower Nc values had higher CBI values. Those gene sequences showing low CBI values suggest the possibility that they are influenced by mutational pressure [17].

Correspondence analysis

Correspondence analysis of the relative synonymous codon usage (RSCU) of all protein coding genes in *Xanthomonas* strains was performed. Correspondence analysis of genes encoding ribosomal proteins and three categories of pathogenicity-related protein genes are shown in Fig. 2.

The first axis strongly correlated with GC3s, whereas axis 2 correlated with G3s. Scatter plot using the first two axes showed genes were focused within a similar region and had very few outliners. Our results with these *Xanthomonas* genomes had a different shape of distribution compared to previous

Table 2
Mean values of Nc, GC, GC3, CAI, CBI and Fop for the genes in six *Xanthomonas* strains

Organism	Genes	Mean Nc	Mean GC%	Mean GC3%	Mean CAI	Mean CBI	Mean Fop
XAC	PCG	37.95	64.81	78.68	0.583	0.293	0.582
	RPG	33.64	61.94	79.12	0.714	0.348	0.621
	GUM	38.43	62.61	78.09	0.560	0.279	0.566
	HRP	43.24	62.27	74.07	0.518	0.211	0.541
	VIR	44.09	60.36	71.53	0.479	0.184	0.523
XCC1	PCG	36.80	65.31	79.95	0.611	0.309	0.592
	RPG	33.88	61.84	79.17	0.725	0.354	0.625
	GUM	39.61	62.91	77.98	0.558	0.274	0.564
	HRP	41.95	63.07	75.54	0.544	0.255	0.563
	VIR	45.34	59.00	68.88	0.474	0.197	0.532
XCC2	PCG	36.54	65.45	80.31	0.620	0.313	0.594
	RPG	33.88	61.84	79.17	0.725	0.354	0.625
	GUM	39.59	62.92	78.00	0.558	0.274	0.564
	HRP	41.96	63.10	75.63	0.544	0.256	0.563
	VIR	45.23	58.76	68.66	0.476	0.192	0.530
XCV	PCG	38.3	64.61	78.10	0.571	0.287	0.579
	RPG	33.40	61.81	79.16	0.716	0.352	0.624
	GUM	38.36	62.78	78.32	0.554	0.276	0.565
	HRP	39.98	63.24	76.31	0.541	0.249	0.563
	VIR	43.29	59.16	71.95	0.490	0.204	0.536
XOO1	PCG	39.12	64.00	77.07	0.583	0.281	0.575
	RPG	34.79	61.38	77.75	0.709	0.326	0.609
	GUM	40.97	60.88	75.23	0.545	0.242	0.544
	HRP	43.84	62.26	73.85	0.520	0.216	0.537
	VIR	38.82	65.30	78.01	0.570	0.264	0.561
XOO2	PCG	39.03	63.92	77.09	0.579	0.281	0.575
	RPG	34.44	61.27	77.61	0.717	0.343	0.618
	GUM	40.67	60.93	75.44	0.542	0.248	0.547
	HRP	43.58	62.93	73.58	0.525	0.224	0.546
	VIR	37.30	65.81	79.59	0.575	0.262	0.559

PCG = protein coding genes; RPG = ribosomal protein genes; GUM = *gum* gene cluster; HRP = *hrp* gene cluster; VIR = avirulence/virulence genes.

studies on genes from *Streptococcus pneumoniae*, *Escherichia coli* or *Nocardia farcinica*, which showed a core region with two ascending horns [28–30]. With all the *Xanthomonas* genomes, the genes were clustered tightly in a core and confined mostly in a narrow range of -0.5 to $+0.5$ of both the axes. No visible ascending masses or horns of genes were observed, but a few genes were found located away from the core. We calculated that on an average 99.43% of the protein coding genes were located in the core block. Genes located away from this core block included a number of hypothetical protein genes, ribosomal protein genes, YapH protein, histone H1 genes, and translation initiation factors (IF1, IF3, etc.). Only one gene associated with pathogenesis (*HrpE*) was found in this region for the XAC genome. These results imply that these six *Xanthomonas* genomes are relatively conserved with most genes lying within the core region. However, some recent reports based on G+C content and CAI values [31] and analysis of nucleotide templates of individual genes with clustering of horizontally transferred genes [32] predicted that a number of pathogenicity related genes acquired through lateral gene transfer. Further study is required to confirm this conclusion.

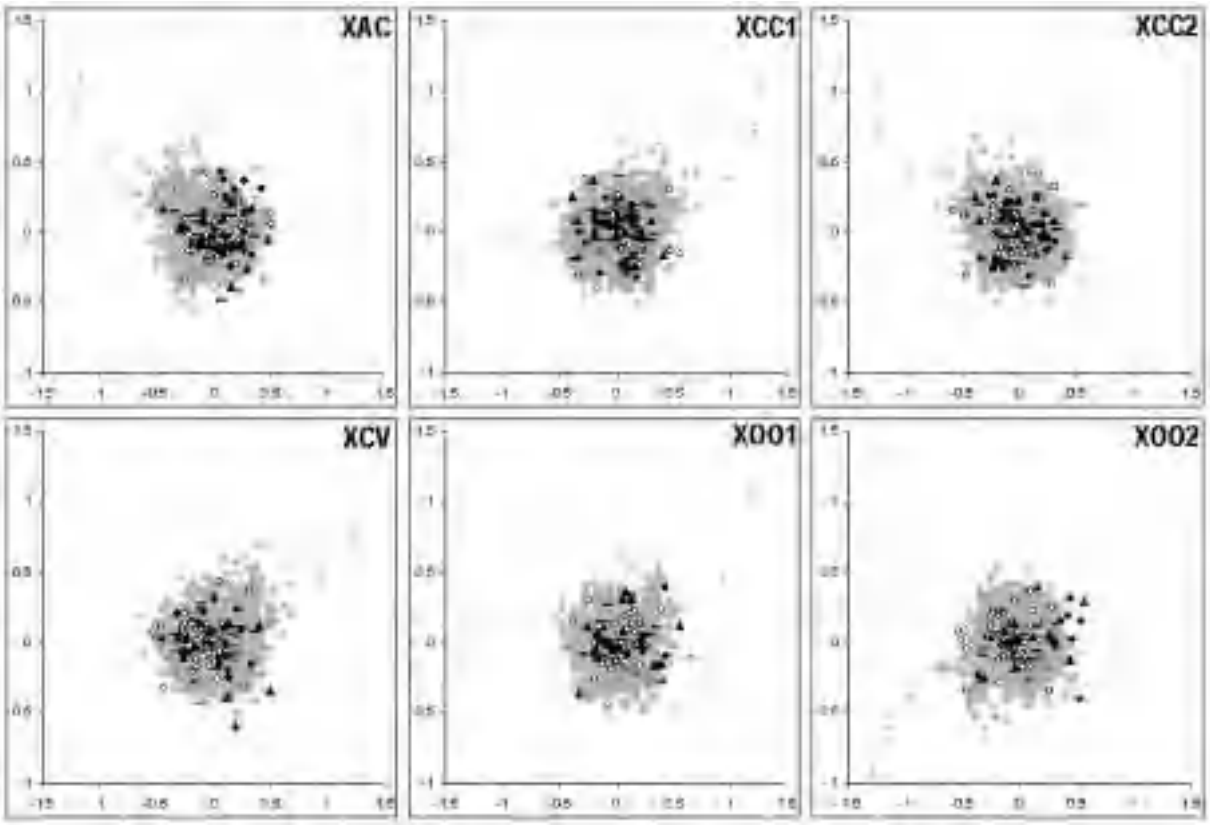

Fig. 2. Correspondence analysis of codon usage patterns of various *Xanthomonas* genomes. In all the plots x and y axis correspond to axis 1 and 2 of the analysis. Buttons are as per Fig. 1.

Identification of potentially highly expressed (*PHX*) *genes in Xanthomonas genomes*

Codon adaptation index (CAI) is a measure of directional synonymous codon usage bias. The index uses a reference set of highly expressed genes from a species to assess the relative merits of each codon, and a score for a gene is calculated from the frequency of use of all codons in that gene. The index assesses the extent to which selection has been effective in molding the pattern of codon usage. In that respect, this index is useful for predicting the level of expression of a gene [12]. Wu *et al.* [26] analyzed proteome results and validated the correlation between CAI values and expression levels showing experimentally that CAI predicted highly expressed genes highly expressed. Thus, we examined CAI values for these *Xanthomonas* genomes to identify the potentially highly expressed genes (PHX genes).

The CAI values for all genes in different *Xanthomonas* strains were calculated and their distributions are shown in Fig. 3.

The CAI values ranged from 0.103 to 0.870, 0.104 to 0.911, 0.104 to 0.911, 0.102 to 0.882, 0.252 to 0.851, and 0.190 to 0.861 in XAC, XCC1, XCC2, XCV, XOO1 and XOO2, respectively. The majority of the genes have CAI values between 0.4 and 0.8. The median CAI values for genes located in the chromosomes were 0.60, 0.63, 0.63 0.59, 0.59 and 0.58 for the aforesaid order of genomes. Plasmid-borne genes of strain XAC showed a CAI value ranging from 0.233 to 0.729 and 0.218 to 0.729 for plasmids pXAC64 and pXAC33 respectively, while strain XCV showed CAI ranges of 0.174 to 0.745,

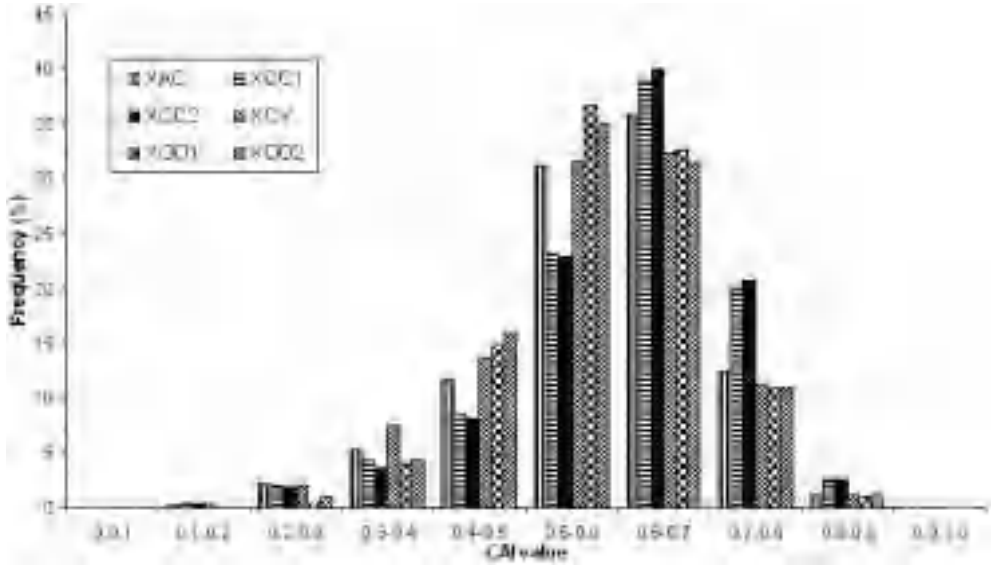

Fig. 3. Frequency distribution of CAI values for all coding genes in the *Xanthomonas* genomes.

0.288 to 0.707, 0.177 to 0.542 and 0.177 to 0.313 for plasmids pXCV183, pXCV38, pXCV19, and pXCV2 respectively. The median CAI values of these plasmid-borne genes were lower than the median value of the chromosomal genes. The median CAI of pXAC64 and pXAC33 were 0.46 and 0.45, respectively, while strain XAC chromosomal genes median CAI was 0.60. Strain XCV chromosomal genes had median CAI value of 0.59 and its plasmids pXCV183, pXCV38, pXCV19 and pXCV2 were 0.43, 0.51, 0.44 and 0.24, respectively.

The average CAI values for different gene groups associated with diverse functions varied (Table 2). Ribosomal protein genes showed high CAI values indicating high levels of gene expression. The predicted expression levels of the virulence genes for strains XOO1 and XOO2 were comparatively higher than those found for the other four species.

As defined by Wu *et al*. [26], the top 10% of the genes, in terms of CAI values, were defined as the predicted highly expressed (PHX) genes. This definition corresponded to CAI cutoffs of 0.716, 0.746, 0.748, 0.713, 0.708 and 0.711 in strains XAC, XCC1, XCC2, XCV, XOO1 and XOO2, respectively. Strain XAC had 445 PHX genes including 30 ribosomal protein genes (RPG) and each plasmid contained one PHX gene. Both plasmid-borne genes were partition protein genes. Strains XCC1, XCC2 and XCV had 25, 22 and 32 RPG in their pool of 433, 466 and 483 PHX genes, respectively. Strain XCV had three PHX genes, which were associated with pXCV183, including one single-stranded DNA-binding protein and two hypothetical protein genes. Strains XOO1 and XOO2 had 22 and 31 RPG among the identified 415 and 441 PHX genes, respectively. Genes associated with pathogenesis genes with two exceptions were found to be expressed at a moderate level and all the other genes were not in the category of PHX. The two exceptions were found in strains XOO1 and XOO2 and both the genes encode a virulence regulator protein. Frequency distribution of CAI values for all coding genes in the genomes of *Xanthomonas* (Fig. 3) showed that maximum numbers of genes were present in the CAI range of 0.6–0.7, but strains XCC1 and XCC2 had statistically significant numbers of genes present in 0.7–0.8 range. These results indicate that strains XCC1 and XCC2 have more highly expressed genes in their genomes than others four strains.

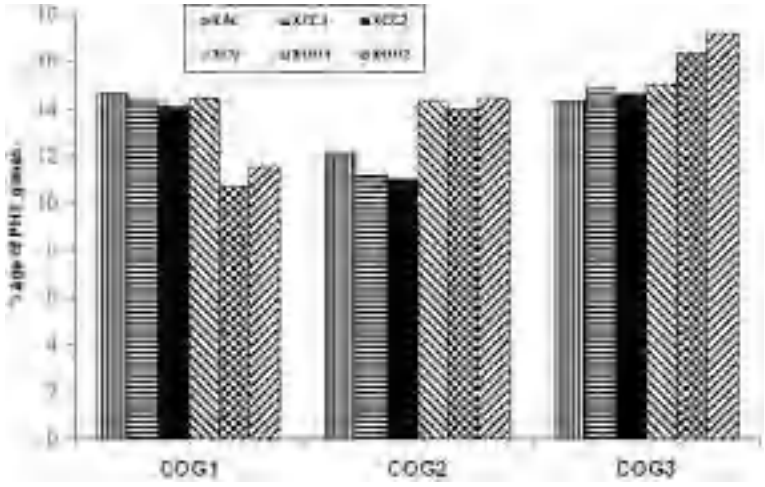

Fig. 4. Percentage of PHX genes in various COG functional groups of *Xanthomonas* genomes.

Clusters of Orthologous Groups of proteins (COGs) were delineated by comparing protein sequences encoded in complete genomes, representing major phylogenetic lineages. Each COG type consists of individual proteins or groups of paralogs from at least 3 lineages and thus, corresponds to an ancient conserved domain. For these *Xanthomonas* genomes, 22 to 23 COG categories were identified and the number of genes found in COG categories are shown in Table 1. To help the analysis, each of the COG categories were clustered in the following 4 COG functional groups: Information storage and processing consisting of COGs related to transcriptions, translation, RNA processing DNA replication and chromatin structure (group 1); cellular processes including, cell division, nuclear structure, defense mechanisms, signal transduction, cell envelop biogenesis, cell motility, cytoskeleton, extra cellular structures, intercellular trafficking and posttranslational modification (group 2); metabolism containing energy production and conversion, carbohydrate transport, amino acid transport, nucleotide transport, coenzyme metabolism, inorganic ion transport and secondary metabolites biosynthesis (group 3); genes with general function predictions and unknown functions (group 4). For this study, we analyzed genes from the first three groups to identify potentially highly expressed (PHX) genes. CAI values of all the genes present in different COG groups were calculated and the PHX genes were identified as per the cut off values of various *Xanthomonas* genomes mentioned above. Figure 4 shows the percentage of PHX genes in various COG functional groups.

It has been found that the percentage of PHX genes in all categories and genomes were above the expected values of 10%. This result implies that the genes in these COG categories are relatively better expressed than the rest of the genes in the genomes. Functional analysis showed that COG functional group 3 (metabolism) contained largest number of PHX genes in all the genomes except in strain XAC where group 1 (information storage and processing) was the largest group. Top 20 PHX genes of all the *Xanthomonas* genomes contain a large number of genes from COG functional categories. Among them, COG group 1 had a major cold shock protein gene and elongation factor P, which is involved in peptide bond synthesis and stimulates efficient translation on native or reconstituted 70S ribosomes *in vitro* [4]; COG group 2 had peptidyl-prolyl *cis-trans* isomerase, an enzyme that accelerates protein folding by catalyzing the *cis-trans* isomerization of proline imidic peptide bonds in oligopeptides [33]; GTP-binding elongation factor protein, promotes the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome [34]; co-chaperonin GroES, binds to Cpn60 in the

presence of Mg-ATP and suppresses the ATPase activity of the latter; molecular chaperone DnaK takes part in bacterial nascent polypeptide chain-folded protein anabolism [35] and chaperonin GroEL, this double ring chaperonin binds unfolded protein, ATP and GroES to the same ring, generating the *cis* ternary complex in which folding occurs within the cavity capped by GroES (*cis* folding) [36] and COG group 3 contained ATP synthase subunit A and B, having hydrolase activity, act on acid anhydrides, catalyzing transmembrane movement of substances [4]; L-glutamine synthetase; spermidine synthase takes part in spermidine biosynthesis and dihydrolipoamide dehydrogenase takes part in alanine and aspartate metabolism, glycine, serine and threonine metabolism, pyruvate metabolism [3].

For the plasmid-borne genes of strain XAC, both pXAC64 and pXAC33 had the same single PHX gene, the partition protein gene, which produces proteins of part A protein family involved in chromosome partitioning [6]. With the plasmids of XCV, three genes on plasmid pXCV183 were PHX genes including two hypothetical protein genes and a short gene that codes for a single-stranded DNA binding protein. Genes from other plasmids were not in the PHX category.

CONCLUSION

The codon usage-based strategy has been successfully applied to the identification of highly expressed genes in various bacteria including G+C rich *Streptomyces* and pathogenic *Nocardia* [26,30,37,38]. In this study, the approach was used to estimate gene expressivity in six strains of phytopathogenic bacteria *Xanthomonas*. The results from this study indicate considerable heterogeneity exists among the genes of these moderately G+C rich genomes. Mostly genes predicted to be expressed highly were house keeping genes and only two out of 339 genes associated with pathogenesis were in the PHX category. Genes present in the plasmids of two strains did not contain many PHX genes. Another important finding was high degree of conservation for all of the genes. Correspondence analysis of these genomes showed that their genes clustered in a very narrow range in both the axes with a few genes outside the core block suggesting that lateral gene transfer was limited in these genomes. These results provided an estimation of the global gene expression patterns in *Xanthomonas* that will be useful in guiding experimental design for further investigation and will enhance our knowledge of the metabolism and pathogenicity of the *Xanthomonas* species.

ACKNOWLEDGEMENT

The work was partially carried out by the grant provided by DBT-India to establish Bioinformatics Facility at NBU. A. Sen acknowledges the receipt of DBT Overseas associateship.

REFERENCES

[1] Ochiai, H., Inone, Y., Takeya, M., Sasaki, A. and Kaku, H. (2005). Genome Sequence of Xanthomonas oryzae pv. oryzae suggests contribution of large numbers of effector genes and insertion sequences to its race diversity. JARQ **39**, 275-287.

[2] Hartung, J. S., Daniel, J. F. and Pruvost, O. P. (1993). Detection of *Xanthomonas* campestnis pv. citri by the Polymerase Chain Reaction Method. Appl. Environ. Microbiol. **59**, 1143-1148.

[3] Lee, B.-M., Park, Y.-J., Park, D.-S., Kang, H.-W., Kim, J.-G., Song, E.-S., Park, I.-C., Yoon, U.-H., Hahn, J.-H., Koo, B.-S., Lee, G.-B., Kim, H., Park, H.-S., Yoon, K.-O., Kim, J.-H., Jung, C.-h., Koh, N.-H., Seo, J.-S. and Go, S.-J. (2005). The genome sequence of *Xanthomonas oryzae* pathovar *oryzae* KACC10331, the bacterial blight pathogen of rice. Nucleic Acids Res. **33**, 577-586.

[4] da Silva, A. C. R., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., *et al*. (2002). Comparison of genomes of two *Xanthomonas* pathogens with differing host specificities. Nature **417**, 459-463.
[5] Qian, W., *et al*. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. Genome Res. **15**, 757-767.
[6] Thieme, F., *et al*. (2005). Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. vesicatoria revealed by the complete genome sequence. J. Bacteriol. **187**, 7254-7266.
[7] Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**, r43-r74.
[8] Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H. and Wright, F. (1988). Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. Nucleic Acids Res. **16**, 8207-8211.
[9] Karlin, S., Campbell, A. M. and Mrazek, J. (1998). Comparative DNA analysis across diverse genomes. Annu. Rev. Genet. **32**, 185-225.
[10] Knight, R. D., Freeland, S. J. and Landweber, L. F. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol. **2**, research0010.
[11] Sharp, P. M. and Li, W. H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. **24**, 28-38.
[12] Sharp, P. M. and Li, W. H. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**, 1281-1295.
[13] Ikemura, T. (1981). Correlation between abundance of *Escherichia coli* tRNAs and their occurrence of the respective codons in protein genes: a proposal for a synonymous codon choice that is optimal for the E coli system. J. Mol. Biol. **146**, 1-21.
[14] Lafay, B., Atherton, J. C. and Sharp, P. M. (2000). Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori*. Microbiology **146**, 851-860.
[15] Dos Reis, M., Wernisch, L. and Savva, R. (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. Nucleic Acids Res. **31**, 6976-6985.
[16] Banerjee, T., Basak, S., Gupta, S. K. and Ghosh, T. C. (2004). Evolutionary forces in shaping the codon and amino acid usages in *Blochmannia floridanus*. J. Biomol. Struct. Dyn. **22**, 13-23.
[17] Sur, S., Sen, A. and Bothra, A. K. (2007). Mutational drift prevails over translational efficiency in Frankia nif operons. Ind. J. Biotechnol. **6**, 321-328.
[18] Peden, J. (1999). Analysis of codon usage. PhD thesis, The University of Nottingham, UK.
[19] Lindgren, P. B. (1997). The role of hrp genes during plant-bacterial interactions. Annu. Rev. Phytopathol. **35**, 129-152.
[20] Rossier, O., Van den Ackerveken, G. and Bonas, U. (2000). HrpB2 and HrpF from *Xanthomonas* are type III-secreted proteins and essential for pathogenicity and recognition by the host plant. Mol. Microbiol. **38**, 828-838.
[21] Dharmapuri, S. and Sonti, R. V. (1999). A transposon insertion in gumG homologue of *Xanthomonas oryzae* pv. *oryzae* causes loss of extracellular polysaccharide production and virulence. FEMS Microbiol. Lett. **179**, 53-59.
[22] Sur, S., Sen, A. and Bothra, A. K. (2006). Codon usage profiling and analysis of intergenic association of Frankia EuIK1 nif genes. Ind. J. Microbiol. **46**, 363-369.
[23] Wright, F. (1990). The "effective number of codons" used in a gene. Gene **87**, 23-29.
[24] Bennetzen, J. L. and Hall, B. D. (1982). Codon selection in yeast. J. Biol. Chem. **257**, 3026-3031.
[25] Ikemura, T. (1985). Codon usage and transfer-RNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**, 13-34.
[26] Wu, G., Culley, D. E. and Zhang, W. (2005). Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. Microbiology **151**, 2175-2187.
[27] Cutter, A. D., Payseur, B. A., Salcedo, T., Estes, A. M., Good, J. M., Wood, E., Hartl, T., Maughan, H., Strempel, J., Wang, B., Bryan, A. C. and Dellos, M.(2003). Molecular correlates of genes exhibiting RNAi phenotypes in *Caenorhabditis elegans*. Genome Res. **13**, 2651-2657.
[28] Médigue, C., Rouxel, T., Vigier, P., Hénaut, A. and Danchin, A. (1991). Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol. **222**, 851-856.
[29] Martín-Galiano, A. J., Wells, J. M. and de la Campa, A. G. (2004). Relationship between codon biased genes, microarray expression values and physiological characteristics of *Streptococcus pneumoniae*. Microbiology **150**, 2313-2325.
[30] Wu, G., Nie, L. and Zhang, W. (2006). Predicted highly expressed genes in Nocardia farcinica and the implication for its primary metabolism and nocardial virulence. Antonie van Leeuwenhoek **89**, 135-146.
[31] Kalia, V., Lal, S. and Cheema, S. (2006). Insight in to the phylogeny of polyhydroxyalkanoate biosynthesis: Horizontal gene transfer. Gene **389**, 19-26.
[32] Tsirigos, A. and Rigoutsos, I. (2005). A Sensitive, support-vector-machine method for the detection of horizontal gene transfers in viral, archaeal and bacterial genomes. Nucleic Acids Res. **33**, 3699-3707.

[33] Fischer, G. and Schmid, F. X. (1990). The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell. Biochemistry **29**, 2205-2212.
[34] Weijland, A. and Parmeggiani, A, (1993). Toward a model for the interaction between elongation factor Tu and the ribosome. Science **259**, 1311-1314.
[35] Bikker, F., Kaman-van Zanten, W. E., de Vries-van de Ruit, A.-M. B. C., Voskamp-Visser, I., van Hooft, P. A. V., Mars-Groenendijk, R. H., de Visser, P. C. and Noort, D. (2006). Evaluation of the antibacterial spectrum of drosocin analogues. Chem. Biol. Drug Des. **68**, 148-153.
[36] Motojima, F. and Yoshida, M. (2003). Discrimination of ATP, ADP and AMPPNP by chaperonin GroEL; Hexokinase treatment revealed the exclusive role of ATP. J. Biol. Chem. **278**, 26648-26654.
[37] Karlin, S. and Mrazek, J. (2000). Predicted highly expressed genes of diverse prokaryotic genomes. J. Bacteriol. **182**, 5238-5250.
[38] Karlin, S., Mrazek, J., Campbell, A. and Kaiser, D. (2001). Characterizations of highly expressed genes of four fast-growing bacteria. J. Bacteriol. **183**, 5025-5040.

7 Morphology, Diversity and Importance of Actinobacteria

Saroja Chhettri*, Gargi Sen Arnab Sen*****

**Assistant Professor, Department of Microbiology, Midnapore College, Midnapore, W.B.*

***Bioinformatics Facility, University of North Bengal, Raja Rammohanpur, Siliguri, West Bengal*

****Department of Botany, University of North Bengal, Raja Rammohanpur, Siliguri, Midnapore, West Bengal*

Abstract

Actinobacteria is a gram-positive bacteria with high content of guanine and cytosine in their DNA. They share the characteristics of both fungi and bacteria. They are prokaryotic like bacteria and are filamentous like fungi. However, they produce more slender nonseptate mycelium. Classic actinobacteria are morphologically characterized by the formation of aerial and substrate mycelium. They are distributed widely in terrestrial and aquatic environments, even exist in some extreme conditions such as thermophilic, psychrophilic, acidophilic and alkaliphilic. Some are found associated with plants, animals and humans. They play important role in biogeochemical cycles and decompose organic matter from dead animals, plants and fungi. Actinobacteria is the source of diverse bioactive secondary metabolites that are important medically, agriculturally and industrially. Actinobacteria especially Streptomyces are exploited for potential applications. This chapter summarizes the morphological characteristics, diversity of actinobacteria in various habitats and their importance.

Keywords: *Actinobacteria, Morphology, Habitats, Importance*

INTRODUCTION

Actinobacteria are gram +ve group of bacteria with high GC content, found in various ecosystems on earth. They share characteristics of both fungi and bacteria. The type of habitat can greatly affect actinobacterial diversity. They are found in many diverse habitat including terrestrial, aquatic, aerial, and even in extreme environments such as *volcanic caves, marine sediments, desserts, hot springs, sediments of cold springs, lakes, etc.*[15] Some are also found in association with higher organisms *such as the genera Frankia present as the symbionts within the root nodules of many woody plants which are dicotyledonous.* [10] The actinobacterial diversity in various habitats can be studied by culture-dependent cultivation and high throughput sequencing techniques. Combining both the techniques is very helpful for the recovery of both the abundant as well as rare members of the microbial community even in extreme environments which were difficult to access [15].

They decompose organic materials such as dead animals and plants debris playing role in humus formation. They play a very important role in carbon cycling as they are known to decompose cellulose, the most abundant biopolymer in nature, followed by chitin.[1] Apart from carbon cycling they are also known for cycling phosphorous, nitrogen, potassium, and many other elements in soil and replenish the nutrient supply in the soil.[25] Geosmin and 2-Methylisoborneal are organic volatile compounds produced by actinobacteria that cause musty and earthy odours in water bodies. Geosmin is responsible for the slightly metallic scent of freshly turned soil. Actinobacteria such as *Streptomyces albidflavus, S. luridiscabei, Nocardia cummidelens* are gross producers of geosmin.[5]

They are known for producing various secondary metabolites such as antibiotics which find its great importance in medicine which includes antifungal, antibacterial and some antitumor drugs. [25] Actinobacteria mainly found in soil especially in the plant rhizosphere have very good plant growth-promoting activity thereby enhancing the growth of plants. They are also known to have antagonistic activity against various plant pathogens causing disease of agriculturally important crops. [13]

This chapter gives a summary of the morphological appearance of actinobacteria, their diversity in various habitats, and their importance.

MORPHOLOGICAL CHARACTERISTICS

Actinobacteria exhibit prokaryotic cell structure with the morphological differentiation highest among gram-positive bacteria. They are branching unicellular microorganisms with most of them being aerobic and characterized by the formation of the aerial and substrate mycelium. [11] Morphologically they appear compact, chalky white, often leathery and covered commonly with aerial mycelium. The substrate mycelium grows either on the culture medium surface or into the medium and is known as the primary mycelium or the vegetative mycelium. Nutrient absorption is the main function of substrate mycelium for actinobacterial growth. They show variation concerning shape, size, thickness and varied colorations which range from white to red, pink, brown, black green, orange or yellow. It is due to adequate differentiation, the thickness of aerial mycelium is frequently greater differentiation compared with substrate mycelium. The structure of aerial mycelium appears as powdery, velvety or cottony, forming concentric zones or rings may have varied colorations due to pigmentation. [1] Some form diffusible pigments. Some of the actinobacteria form spores, sporangia and sporangiophore. The position and the number of spore, ornamentation of spore surface, sporangia shape, sporangiophore with the presence or absence of flagella are some morphological characteristics important in the classification of actinobacteria. [11]

DIVERSITY AND IMPORTANCE OF ACTINOBACTERIA

Terrestrial Habitat

Actinobacteria can be found in both terrestrial and aquatic habitats. They are economically important to humans since the actinobacteria such as *Streptomyces* are the contributors of forest and agricultural soil systems as they decompose dead organic matter. They are also known for the biological buffering of soil. The growth of several pathogens that cause disease to plants is inhibited by actinobacteria in the rhizosphere thereby promoting the production of crops. [3] About 80% of the antibiotics in the world are from actinobacteria especially from the genera *Micromonospora* and *Streptomyces.* Antibiotics such as erythromycin, tetracycline and vancomycin are produced by actinobacteria. These soil dwellers are also the source of many anticancer drugs, insecticides and herbicides. Soil isolates such as *Nocardia bangladeshensis, Streptomyces rimosus, S. iakyrus*, *S. bingchengensis, S thermocarboxydovorans, S. luteospreus, S. alboniger* and *S. gilvosporeus* have antibacterial and antifungal properties. [4]

Aquatic Habitat

The biodiversity present on earth is contributed largely by the aquatic ecosystem. The basic freshwater ecosystem consists of lotic (rivers and streams), lentic (ponds, lakes and pools) and wetlands. Actinobacteria is among the abundant group of microbes found in freshwater habitats. However, maximum actinobacterial strains were found from the lake sediments compared to that from the river sediments. This may be because of the continuous running water in rivers whereas sediments in the lake are not much affected by the running water. Huge actinobacterial diversity was found in freshwater isolates from sediments of lakes and rivers which included common genera *Streptomyces* and eight rare genera included *Rhodococcus Amycolatopsis, Saccharopolyspora, Nocardiopsis, Micrococcus, Kocuria Prauserella, Promicromonospora*. [17] Acquatic actinobacteria have some ecological roles as they degrade some recalcitrant organic matter in the water bodies. They like soil microbes are capable of degrading organic matter from plant biomass. The ability of aquatic actinobacteria to uptake nutrients like phosphorous, nitrogen and carbon enhances their performance under nutrient-depleted or oligotrophic conditions. They also help in preventing eutrophication, as they can sequester inorganic phosphorous in water bodies. [8]

Symbiotic actinobacteria

Actinobacteria interact with both micro and macroorganisms. There is a symbiotic relation between actinobacteria and the hosts such as animals, plants, humans and insects where the actinobacteria protect the pathogens against the host organisms. [2] Actinobacteria such as *Frankia* form a symbiotic association with woody plants mostly from eight families and three orders namely Rosales, Fagales and Cucurbitales. The actinorhizal nodules comprise modified lateral roots with the cells which are infected in the expanded cortex. Nitrogen is present in the atmosphere as the most abundant element, however, most of the plants are not able to utilize the atmospheric nitrogen directly and are dependent on soil for nitrogen sources. The root nodules that are present on the plant root system contain symbiotic nitrogen fixing actinobacteria such as *Frankia* strains which promote the fixation of nitrogen. The oxygen sensitive enzyme nitrogenase is protected by *Frankia* and thus facilitating nitrogen nutrition in plants. [23]The symbiotic association between Frankia and actinorhizal plant belonging to *Casuarinaceae* family such as *C.equisetifolia* and *C. glauca*. Agroforestry widely uses

Casuarina trees for the reclamation of land and is also important for crop protection, windbreaks firewood, source of poles and charcoal. Actinorhizal plants are with ecological and economic benefits including stabilization of soil, reforestation, and reclamation of land. [14]

Endophytic Actinobacteria

Endophytic actinobacteria reside symbiotically within the plant tissues as endophytes. They play a role of importance in the development and growth of the host plant by producing large amount of bioactive natural products. Lately, they are of great interest because them being a good source of some novel compounds which may have applications in agriculture, medicine, and the environment. There is great diversity among the endophytic actinobacteria including those isolated from plants in specific ecological niches and extreme habitats. [22] In recent studies 169 endophytic actinobacteria were isolated from *Rhynchotoechum ellipticum* with 81 strains with antimicrobial potential. Some *Streptomyces* sp. among them including *Streptomyces olivaceus* and *Streptomyces thermocarboxydus* produced antibiotics such as miconazole, rifampicin, fluconazole, ketokonazole and erythromycin. Some of them also produced anticancer compounds such as paclitaxel. The antagonistic actinobacteria *Streptomyces olivaceus* and *Streptomyces thermocarboxydus* showed plant growth promoting traits such as phosphate solubilization. Endophytic actinobacteria belonging to genera such as *Streptomyces, Actinomyces, Micromonospora, Microbacterium, Leifsonia, Pseudonocardia, Kocuria, Amycolatopsis* and *Brevibacterium* are producers of plant growth hormone indole acetic acid. [16]

Actinobacteria Residing in the human body

Actinobacteria is found in diverse habitats including the human body. *Propionibacterium, Actinomyces, Rothia, Bifidobacteria* and *Corynebacterium* are the five important genera of actinobacteria residing in healthy individuals. Some commensals and opportunistic pathogens also belong to this phylum. [24] Actinobacteria is also one of the major phyla among the other phyla that constitute the gut microbes. Actinobacteria such as *Bifidobacteria* sp. are known for health-promoting effects and affect human nutrition and metabolism positively. They constitute the intestinal flora of mammals including animals and can be used as probiotics for the treatment and prevention of diseases.[9]

Extremophilic Actinobacteria

Apart from actinobacteria in temperate habitats, they are also distributed in some extreme habitats of the terrestrial and aquatic ecosystem. Extremophilic actinobacteria such as thermophiles can grow at a temperature range of 40ºC to 80ºC. *Streptomyces thermoautotrophicus*, *Acidithiomicrobium* sp., *Amycolatopsis methanolica* are some of the thermophilic actinobacteria. [21] Psychrophilic actinobacteria such as *Streptomyces, Lentzea, Amycolatopsis* did grow in the temperature range of 10 ºC to 25 ºC with the capability to produce enzymes such as cellulase, amylase, protease and antagonistic activities. Psychrotolerant Actinobacteria belonging to genera *Tsukamurella, Streptacidiphilus, Rhodococcus, Streptomyces, Arthrobacter, Actinoplanes, Nocardia, Kribbella, Pseudarthrobacter* and *Pilimelia* were identified from Antarctica. Some of them were producers of potential antitumor compounds.[20]

Actinobacteria are found to occur in a diverse range of pH with some acidophiles found to occur at lower pH. In the recent studies, it was found a large proportion of acidophilic actinobacteria compared to neutrophiles, exhibited good plant growth-promoting activities and played important role in phosphate solubilization, siderophore production and showed antifungal activities. Those acidophilic actinomycetes beneficial to plants were isolated from rhizospheric soil using the media pH of 5.5 and were identified as *Streptomyces misionensis* and non *Streptomyces* isolates such as *Verrucosispora, Sacchaopolyspora, Nocardia, Mycobacterium, Amycolatopsis, Allokutzneria and Nonomuraea*. [18]

Some Alkaliphilic actinobacteria are found in alkaline soil, alkaline lake, soda lake. Alkalitolerant actinobacteria grow in both alkaline and neutral environments. They grow in the pH range of 7 to 11. *Streptomyces caeruleus*, *S. canescens , S. cavourensi, S. hydrogenans and Nocardiopsis alborubida are some of the alkalophilic actinobacteria.* [21]

Actinobacteria are also adapted to grow in conditions of high salt concentrations. Their ability to adapt themselves to high osmotic pressures existing in the environment makes them salt lovers. They live in different halophilic environments such as Deep sea brines, soda lakes, marine sediments marsh soil with 10% NaCl and many other hypersaline regions *Nocardiopsis halotolerans, Saccharomonospora halophila* and *Streptomonospora alba, Nocardiopsis salina* are some halotolerant and halophilic actinobacteria growing at high salt concentrations. [12] Haloalkane and hypersaline conditions prevalent in the mangroves and lakes limit the ability of the microbes

to hydrolyze biomolecules like lignin, chitin and cellulose. Only haloalkalitolerant and haloalkaliphilic actinobacteria and bacteria can proliferate and decompose the recalcitrant biopolymers. [21]

The presence of genes for stress response, osmoregulation, heavy metal resistance and some genes for antibiotic production and antibiotics in the genome sequence of some *Streptomyces* isolates would perhaps make the survival of actinobacteria possible in extreme environments like halophilic. [6]However their mechanism of adaptation, biological activities and diversity is less studied. The extreme and special environments are the source of some novel species and rare actinobacteria. They are new sources for the exploitation in the field of agriculture, industry and medicine as they may be sources of some novel natural product. [19] *The major genera comprised of Streptomyces and Nocardiopsis among 200 actinobacterial strains isolated from an extreme environment such as Great Salt Plain with high salinity, high temperatures, and exposed to UV radiation.* [7]

CONCLUSION

Actinobacteria is the dominant group of microbes that are distributed in various habitats such as aquatic, terrestrial, and in extreme conditions such as thermophilic, psychrophilic, acidophilic and alkalophilic. Their survival in extreme and special habitats proves their adaptability. It can be inferred from the review that they have many practical applications as they are a source of many secondary metabolites which are medically, agriculturally and industrially important and is of commercial value for the welfare of humanity. However many rare genera of actinobacteria are yet to be explored from many diverse unexplored locations for their industrial and biotechnological potential and they can be of value in the discovery of new drugs.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Anandan, R., Dharumadurai, D., & Manogaran, G. P. (2016). In *Actinobacteria-Basics and Biotechnological Applications*. Intechopen.
2. Anteneh, Y. S., & Franco, C. M. (2017). In *Biology and Biotechnology of Actinobacteria* (pp. 233-268).
3. Bhatti, A. A., Haq, S., & Bhat, R. A. (2017). *Microbial pathogenesis*, *111*, 458-467.

4. Chanthasena, P., & Nantapong, N. (2016). *Brazilian Archives of Biology and Technology*, *59*.
5. Churro, C., Semedo-Aguiar, A. P., Silva, A. D., Pereira-Leal, J. B., & Leite, R. B.(2020). *Scientific reports*, *10*(1), 1-18.
6. Cornell, C. R., Marasini, D., & Fakhr, M. K. (2018). *Frontiers in microbiology*, *9*, 2282.
7. Gad, A. H. (2014). *Bacterial diversity at the great salt plains of Oklahoma*. Doctoral dissertation, The University of Tulsa.
8. Ghai, R., Mizuno, C. M., Picazo, A., Camacho, A., & Rodriguez-Valera, F. (2014).*Molecular ecology*, *23*(24), 6073-6090.
9. Hidalgo-Cantabrana, C., Delgado, S., Ruiz, L., Ruas-Madiedo, P., Sánchez, B., & Margolles, A. (2017). *Microbiology spectrum*, *5*(3), 5-3.
10. Kurtboke, D.I. (2017). In *Biology and biotechnology of Actinobacteria* (pp. 123-149).
11. Li, Q., Chen, X., Jiang, Y., & Jiang, C. (2016). *Actinobacteria-basics and biotechnological applications*, 59-86.
12. Mohamad, O. A. A., Li, L., Ma, J. B., Hatab, S., Rasulov, B. A., Musa, Z., Liu, Y.H. & Li, W. J. (2018). In *Extremophiles in Eurasian Ecosystems: Ecology, Diversity, and Applications* (pp. 333-364).
13. Nalini, M. S., & Prakash, H. S. (2020). *Plant Microbiomes for Sustainable Agriculture*, *25*, 199.
14. Ngom, M., Gray, K., Diagne, N., Oshone, R., Fardoux, J., Gherbi, H., Hocher, V., Svistoonoff, S., Laplaze, L., Tisa, L.S. and Sy, M.O.,(2016). *Frontiers in plant science*, *7*, 1331.
15. Rego, A., Raio, F., Martins, T.P., Ribeiro, H., Sousa, A.G., Séneca, J., Baptista, M.S., Lee, C.K., Cary, S.C., Ramos, V. and Carvalho, M.F., (2019). *Frontiers in microbiology*, *10*, 1018.
16. Passari, A.K., Mishra, V.K., Singh, G., Singh, P., Kumar, B., Gupta, V.K., Sarma, R.K., Saikia, R. and Singh, B.P., (2017). *Scientific reports*, *7*(1), 1-17.
17. Passari, A. K., Leo, V. V., Chandra, P., Kumar, B., Nayak, C., Hashem, A., Abd_Allah, E.F., Alqarawi, A.A. & Singh, B. P. (2018). *Microbial cell factories*, *17*(1), 1-14.
18. Poomthongdee, N., Duangmal, K., & Pathom-aree, W. (2015). *The Journal of antibiotics*, *68*(2), 106-114.
19. Qin, S., Li, W. J., Klenk, H. P., Hozzein, W. N., & Ahmed, I. (2019). *Frontiers in microbiology*, *10*, 944.
20. Silva, L.J., Crevelin, E.J., Souza, D.T., Lacerda-Júnior, G.V., de Oliveira, V.M., Ruiz, A.L.T.G., Rosa, L.H., Moraes, L.A.B. and Melo, I.S., (2020). *Scientific reports*, *10*(1), 1-15.

21. Shivlata, L., & Tulasi, S. (2015). *Frontiers in microbiology*, *6*, 1014.
22. Singh, R., & Dubey, A. K. (2018). *Frontiers in microbiology*, *9*, 1767.
23. Van Nguyen, T., & Pawlowski, K. (2017). In *Rhizotrophs: Plant Growth Promotion to Bioremediation* (pp. 237-261).
24. Wu C. (2013). In: Nelson K. (eds).Encyclopedia of Metagenomics.
25. Zhang, B., Wu, X., Tai, X., Sun, L., Wu, M., Zhang, W., Chen, X., Zhang, G., Chen, T., Liu, G. and Dyson, P., (2019). *Frontiers in microbiology*, *10*, 2209.

Indian J. Applied & Pure Bio. Special Volume 140-146 (2022).
Environment & Sustainable Development : Problems, Prospects & Mitigation Date : 25th-27th November, 2021
ISSN: 0970-2091
A web of Science Journal

Isolation and identification of *Streptomyces* sp. producing agroactive enzymes with biocontrol potential

Saroja Chhettri[1,2], Gargi Sen[3] and Arnab Sen[1,2,3]*

[1]Midnapore College, Midnapore, West Bengal 721101 (India)
[2]Department of Botany, University of North Bengal, Raja Rammohanpur, Siliguri-734013 (India)
[3]Bioinformatics Facility, University of North Bengal, Raja Rammohanpur, Siliguri-734013 (India)
***Corresponding Author:** arnab.sen@nbu.ac.in

Abstract

The enzyme chitinase has potential application in the field of agriculture. The aim of the present study is the isolation, characterization and identification of the potential chitinase-producing bacteria from the Nagari farm tea garden of Darjeeling Hills. Around twelve actinobacterial strains were isolated and screened for their chitinase and cellulase activities followed by the antagonistic activity against fungal pathogen *Fusarium* sp. The potential isolate with antagonistic activity was then characterized and identified as *Streptomyces* sp. SDr 22. The isolate tolerated the NaCl concentration of 1% to 4% and could grow in the pH range of 5 to10 at an optimum temperature of 30ºC. The presence of chitinase-producing *Streptomyces* isolate in the tea garden of Darjeeling Hills with biocontrol activity has been explored in this study.

There is a great demand for agricultural land for the cultivation of crops to supply food to the ever-increasing world population. However, there are limitations of the land that can be cultivated which increases the immediate need to control crop diseases and increase the yield with the limited land resource. Both the abiotic and biotic factors are responsible for the productivity of crop plants. One of the major factors causing a decrease in the yield of crops is fungal plant pathogens.[12] The pathogens causing various diseases to plants are being controlled by various chemical insecticides and fungicides. However, these chemicals are also causing harm to the environment thereby affecting both animal and human health.[33] Global warming accompanied by climatic change has increased the incidence of various fungal diseases and causing damage to crop plants. This has posed a great loss in the agricultural sector and production of food.[2]

Chitin is present in the exoskeleton of arthropods such as crustaceans and insects and is the most abundant biopolymer in nature

after cellulose.[33] It is also one of the components of the cell wall of fungal pathogens. Chitinases play a role of importance in the degradation of chitinous waste.[35] Chitinases are glycosyl hydrolases that hydrolyze the chitin in the insoluble form to pharmaceutically valuable products such as chitooligosaccharides and glucosamines which are soluble forms. These chitooligosaccharides are known to have antifungal, antibacterial, antitumor and immune-enhancing effects. The bacteria producing chitinases with biocontrol potential against pathogenic fungi have been exploited.[17] Thus the role of chitinase producing microbes as biocontrol agents cannot be ignored in such a scenario.

Actinobacteria is one of the main sources of bioactive secondary metabolites and commercially important enzymes in both the medical and agricultural fields.[16] They play important role in the cycling of carbon sources such as chitin and cellulose.[14] Actinobacteria especially from the genus *Streptomyces* which are are filamentous, gram-positive, long rods have chitinolytic activity. Several *Streptomyces* sp such as *S. griseus, S. lividans, S. plicatus, S.aureofaciens, S. halstedii, S.glauciniger* are known for producing chitinolytic enzymes.[4,16,30] They have antagonistic activity towards many pathogenic fungal species. *Streptomyces* is known to control Fusarium wilt.[37] The considerable affinity of chitinase to chitin has lead several biotech companies to explore the potential for the development of disease-resistant seeds and transgenic plants.[12]

In the present study, we isolated the actinobacterial strains from Nagari farm tea garden soil. The strains recovered were screened for the production of the enzymes cellulase and chitinase. The potential enzyme producer selected and studied further for their antifungal activities against the phytopathogen *Fusarium* sp. This isolate with the good antagonistic activity was characterized morphologically, biochemically and physiologically and identified by 16s rDNA sequencing.

Isolation :

Tea rhizospheric soil samples were collected from Nagari farm tea garden located in Darjeeling hills, West Bengal, India. The soil samples collected were air-dried for five days at 30ºC. The soil sample dried and was serially diluted in saline water containing 0.85% NaCl.[31] The inoculation was done from each dilution by the method of spread plate. For isolation, the media used for growth were Starch nitrate agar[11] and Inorganic Salts, Starch Agar (ISP4).[31] The inoculated plates were incubated at 30°C for seven days. After purification, the isolates were sub-cultured using ISP4 slants and stored at 4°C. [24].

Screening for the production of chitinase and cellulase :
Chitinase assay :

The isolated strains inoculated into chitin agar media (Colloidal chitin 5g; KH_2PO_4 0.03g; K_2HPO_4 0.07g; $FeSO_4.7H_2O$ 0.001g; $ZnSO_4.4H_2O$ 0.0001g; $MgSO_4.7H2O$ 0.05g; $MnCl_2.4H2O$ 0.0001g; agar 2g; ddH_2O 100 ml, pH adjusted to 7.0). Colloidal chitin was prepared[7] and the plates after inoculation were incubated for 5 days and the plate flooded with gram's iodine solution and the halo around the colonies were recorded.[22]

Cellulase production :

The isolated strains were inoculated into CMC Agar (Carboxy-methylcellulose 0.5g, K_2HPO_4 0.1 g, $NaNO_3$ 0.1 g, yeast extract 0.05 g, $MgSO_4$ 0.05 g, Agar15 g, distilled water 1000 ml) plates and incubated at 30°C for 5 days. After completion of incubation, plates were flooded with Gram's iodine solution which resulted in the formation of a bluish-black complex with unhydrolyzed cellulose. The clearance zone around the colonies is the indicator of a positive result.[10]

Antagonistic activity against fungal pathogens:

The strains under study were tested in vitro for their antagonistic activity against fungal root pathogen, *Fusarium solani* (RHS/P388) with accession number NAIMCCF-02901 and other *Fusarium* species obtained from immunopathology Laboratory, Department of Botany, North Bengal University. The antagonism test was carried out in Potato Dextrose Agar (PDA) by the method of dual culture[36] with slight modification. Freshly grown *Fusarium* sp. inoculated on one edge of the plate and the isolated strain SDr 22 inoculated in the same plate at the extreme opposite edge of the plates by streaking, sealed tightly with parafilm. The control plate was only inoculated with *Fusarium* sp. After inoculation plates were incubated for 10 days at 30°C.

Morphological studies :

Morphological studies were done by growing the isolates at 30°C for 7 days in various International Streptomyces Project (ISP) Medium such as Tryptone-yeast extract broth[19], Yeast extract-malt extract agar,[20] Oatmeal agar,[13] Glycerol-asparagine agar,[21] Inorganic salts-starch agar[13] Peptone-yeast extract iron agar,[32] Tyrosine agar.[25] The formation of the substrate and aerial mycelium was studied.[26]

Biochemical and Physiological Studies :

The biochemical tests that were carried out were the formation of Melanin,[26] degradation of Casein Tyrosine, Xanthine,[5] hydrolysis of Esculin[9], production of Amylase,[8] Gelatinase,[27] Lipase,[3] reduction of Nitrate,[15] and Carbohydrate utilization test.[26] Growth of the isolate in various pH(5-10),NaCl concentration (1% to 7%) and various temperature (25°C,30°C,40°C) evaluated using Bennet Agar media.[28]

DNA isolation and 16srDNA sequencing :

The isolation and purification of DNA of the potential isolate were carried out following CTAB method.[34] The amplification of 16S rDNA sequence carried out by Polymerase Chain Reaction (PCR) using primers ACT235F (5'CGC GGC CTA TCA GCT TGT TG3') and ACT 878R (3'CCG TAC TCC CCA GGC GGG G5'.[29]

Isolation and Screening for the production of chitinase and cellulase :

Colonies of actinobacteria were found to be white and chalky in the isolation media used namely Inorganic Salts Agar medium and Starch Nitrate agar medium. Total twelve

colonies of actinobacteria were recovered and they were found to be a producer of chitinase and cellulase. The isolate SDr 22 with the most potential ability to produce cellulase (Fig 1a) and chitinase (Fig 1b) were also found to exhibit good antifungal activity against plant pathogenic fungus *Fusarium* sp (Fig 1c). The control plate was fully grown by the *Fusarium* sp. Whereas in the plate with both *Fusarium* sp and SDr 22 inoculated, the growth and spread of *Fusarium* sp. in the entire plate was suppressed.

Morphological, biochemical, physiological characterization and 16s rDNA sequencimg:

Morphological studies of the isolate showed the strain showed variation in the coloration of Aerial and Substrate mycelium in all the tested media with variation of spore mass coloration (Table-1).

In biochemical studies it was found that the isolate produced amylase, cellulase and lipase and utilized tyrosine and esculin however failed to produce gelatinase, reduce nitrogen and hydrolyze casein. In carbohydrate utilization studies it was found the isolate SDr 22 utilized all the tested sugars namely glucose, arabinose, fructose, sucrose, xylose, inositol, rhamnose and raffinose. The isolate tolerated NaCl concentration of 1% to 4% and showed growth in the pH range of 5 to 8.

The Molecular identification of the strain was done by Sanger sequencing of 16S rDNA gene fragment after amplification by PCR and the phylogenetic tree constructed by the neighbor-joining method with the sequence obtained from *Streptomyces* sp SDr 22 and other *Streptomyces* sp.(Fig 2).

The similarity of the sequence studied through the online NCBI BLAST program, where it was found the isolate SDr 22 showed 99.50 % similarity to *Streptomyces* sp. By the similarity-based searches in the ezBioCloud server the isolate showed a similarity of 97.25% to *Streptomyces flavovirens*. The 16S rDNA gene sequences were submitted to NCBI and the accession number provided was MK300088.1.

A good percentage of the soil microbial community comprises actinobacteria with *Streptomyces* being the major genus known for producing antibiotics, bioactive compounds and various extracellular enzymes of importance. This genus has huge potential to improve agricultural productivity in the future.[18] Our study supports the findings where the *Streptomyces and* non-*Streptomyces* species producer of hydrolytic enzymes such as cellulase, chitinase, protease, and pectinase could degrade fungal and bacterial cell walls thus causing antagonism against fungal pathogens.[6] The *Streptomyces* sp. with chitinolytic activity plays the role of importance in the biocontrol of plant diseases caused by *Fusarium* sp. and attention is also being given to identifying the bioactive metabolites that are accountable for biocontrol. *Streptomyces* is the major microbe responsible for the suppression of diseases in plants.[23] Plant growth-promoting rhizobacteria including *Streptomyces*, with antagonistic activity against *Fusarium* sp, induces the immune system of the plant against various biotic stresses and thus are the better alternative to chemical pesticides and fungicides in the agricultural system.[1]

Table-1. Morphological characterization of SDr 22 in various ISP media

Medias	Aerial	Substrate mycelium mycelium	Spore Mass	growth
Tryptone Yeast Extract Agar	Off White	Off White	No Spore Mass	++
Yeast Malt Agar	Off White	Off white to Brown	Whitish	+++
Oatmeal Agar	Off White	Brownish Grayish	Brownish	+++
Inorganic Salt Starch Agar	White	Greenish	Olive Green	+++
Glycerol Asparagine Agar	Off White	Off White	Off White	++
Peptone Yeast Extract Iron Agar	Off White	Off white to Light Brown	Off White	+++
Tyrosine Agar	Off White	Grayish Brownish	Grayish	+++

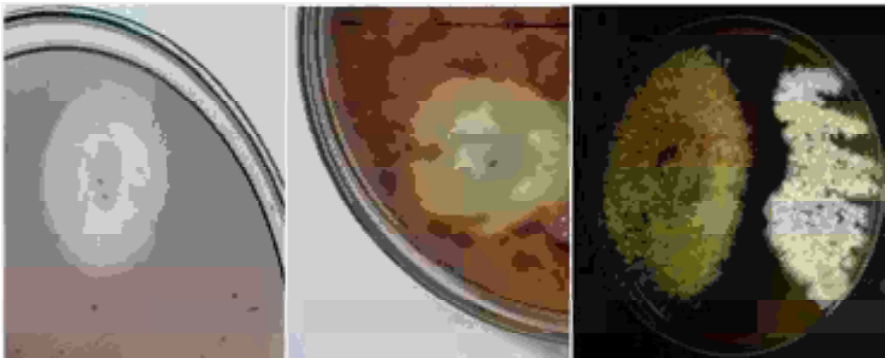

Fig1 (a). Cellulase production by strain SDr 22 in CMC agar after flooding with iodine solution (b). Chitinase production in Colloidal chitin agar by strain SDr 22, with a zone of clearance around colony after flooding with iodine solution. (c). Antagonistic activity of SDr 22 against *Fusarium* sp. after 10 days of incubation.

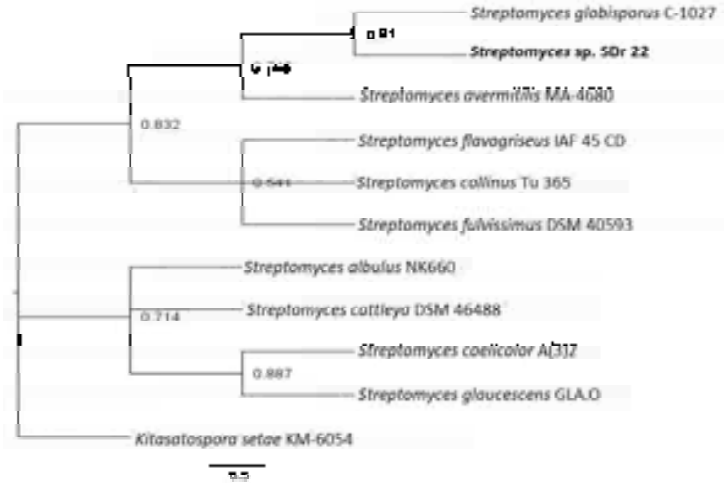

Fig 2: Phylogenetic tree based on the partial 16s rRNA sequences

In this study, the actinobacterial strains from tea garden soil were assessed for the production of cellulase and chitinase. The potential enzyme-producing isolate SDr 22, with good antifungal activity, was found to belong to genus *Streptomyces* based on morphological, biochemical and 16srDNA sequencing studies. Thus *Streptomyces* sp. SDr 22 represents the indigenous tea rhizosphere population of tea rhizospheric soil and can be explored for the preparation of biofertilizer in future. Chemical pesticides and fungicides are being used for many years however the use of eco-friendly alternatives will be a boon to the ecosystem thereby preventing the damage further.

Conflict of Interest :

The authors declare no conflict of interest.

Authors would like to acknowledge Bioinformatics Facility, the Department of Botany, University of North Bengal and Biswa Bangla Genome Centre, University of North Bengal for providing the facilities to carry out the research work.

References:

1. Abbasi, S., N. Safaie, A. Sadeghi and M. Shamsbakhsh (2019). *Frontiers in Microbiology*, *10:* 1505.
2. Almeida, F., M.L. Rodrigues and C. Coelho (2019). *Frontiers in microbiology*, *10:* 214.
3. Aly, M. M., S. Tork, S. M. Al-Garni and L. Nawar (2012). *African journal of Microbiology research*, *6*(6): 1125-1137.
4. Awad, H. M., H. A. El-Enshasy, S. Z. Hanapi, E.R. Hamed and B. Rosidi (2014). *Natural product research*, *28*(24): 2273-2277.
5. Berd, D. (1973). *Applied microbiology*, *25*(4): 665-681.
6. Borah, A. and D. Thakur (2020). *Frontiers in microbiology*, *11*(318):1-23.
7. Deepthi, A. and P. Rosamma (2014). Doctoral dissertation, Cochin University of Science And Technology.
8. Gopinath, S.C., P. Anbu, M.M. Arshad, T. Lakshmipriya, C. H. Voon, U. Hashim and S. V. Chinni (2017). *BioMed research international*, 1-9.
9. Gordon, R.E., D.A. Barnett, J.E. Handerhan, and C. H. N. Pang (1974). *International Journal of Systematic and Evolutionary Microbiology*, *24*(1): 54-63.
10. Kasana, R. C., R. Salwan, H. Dhar, S. Dutt and A. Gulati (2008). *Current microbiology*, *57*(5): 503-507.
11. Korayem, A.S., A. A. Abdelhafez, M. M. Zaki and E. A. Saleh (2015). *Annals of Agricultural Sciences*, *60*(2): 209-217.
12. Kumar, M., A. Brar, M. Yadav, A. Chawade, V. Vivekanand and N. Pareek (2018). *Agriculture*, *8*(7): 88.
13. Kuster, E. (1959). *International Journal of Systematic and Evolutionary Microbiology*, *9*(2): 97-104.
14. Lacombe-Harvey, M. È., R. Brzezinski and C. Beaulieu (2018). *Applied microbiology and biotechnology*, *102*(17): 7219-7230.
15. Maciejewska, M., D. Adam, A. Naome, L. Martinet, E. Tenconi, M. Całusińska, P. Delfosse, M. Hanikenne, D. Baurain,, P. Compere and M. Carnol (2017). *Frontiers in microbiology*, *8*(1181: 1-13.
16. Narayana, K. J. and M. Vijayalakshmi (2009). *Brazilian Journal of Microbiology*, *40*(4): 725-733.

17. Okongo, R. N., A. K. Puri, Z. Wang, S. Singh and K. Permaul (2019). *Journal of Bioscience and Bioengineering*, *127*(6): 663-671.
18. Olanrewaju, O. S. and O. O. Babalola (2019). *Applied microbiology and biotechnology*, *103*(3): 1179-1188.
19. Pridham,T. G. and D. Gottlieb (1948). *Journal of Bacteriology*, *56*(1): 107-114.
20. Pridham, T.G., P. Anderson, C. Foley, L.A. Lindenfelser, C.W. Hesseltine and Benedict (1957). *Antibiotics annual*, 947-53.
21. Pridham, T.G. and A. J. Lyons Jr (1961). *Journal of Bacteriology*, *81*(3): 431-441.
22. Sharma, V. and R. Salwan (2015). *Indian Journal of of Applied Microbiology*, *18:* 1-6.
23. Sharma,V., A. Sharma, A. B. Malannavar and R. Salwan (2020). In *Molecular aspects of plant beneficial microbes in agriculture*. Academic Press. *7:* 89-109.
24. Shepherd, M. D., M. K. Kharel, M. A. Bosserman and J. Rohr (2010). *Current protocols in microbiology*, *18*(1): 10E-1.
25. Shinobu, R. (1958). *B Natural Science*. *7:* l-76.
26. Shirling, E.T. and D. Gottlieb (1966). *International Journal of Systematic and Evolutionary Microbiology*, *16*(3): 313-340.
27. Smith Jr, H. L. and K. Goodner (1958). *Journal of bacteriology*, *76*(6): 662-665.
28. Sreevidya, M., S. Gopalakrishnan, H. Kudapa and R. K. Varshney (2016). *Brazilian journal of microbiology*, *47:* 85-95.
29. Stach, J. E., L. A. Maldonado, A. C. Ward, M. Goodfellow and A. T. Bull (2003). *Environmental microbiology*, *5*(10): 828-841.
30. Suarez-Moreno, Z. R., D. M. Vinchira-Villarraga, D. I. Vergara-Morales, L. Castellanos, F. A. Ramos, C. Guarnaccia, G. Degrassi, V. Venturi and N. Moreno-Sarmiento (2019). *Frontiers in microbiology*, 10: 290.
31. Taddei, A., M. J. Rodríguez, E. Márquez-Vilchez and C. Castelli (2006). *Microbiological Research*, *161*(3): 222-231.
32. Tresner, H. D. and Danga, F. (1958). *Journal of bacteriology*, *76*(3): 239-244.
33. Veliz, E.A., P. Martínez-Hidalgo and A.M. Hirsch (2017). *AIMS microbiology*, *3*(3): 689-705.
34. William, S., H. Feil and A. Copeland (2012). *Sigma*, *50*: 6876.
35. Xie, X. H., X. Fu, X. Y. Yan, W. F. Peng and L. X. Kang (2021). *Marine Drugs*, *19*(7): 356.
36. Zhao, L., Y. Xu and X. Lai (2018). *Brazilian journal of microbiology*, *49:* 269-278.
37. Zou, N., D. Zhou, Y. Chen, P. Lin, Y. Chen, W. Wang, J. Xie and M. Wang (2021). *Frontiers in microbiology*, *12:* 2226.

ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Bioteehnology 8 (4): 434-441, 2009
ISSN 1682-296X

In Silicn Analysis nf Evniutinn in Swine Fln Virai Genomes Through Re-assortment hy Promnlgation and Mntation

[l]S. Sur, [l]G. Sen, [l]S. Thaknr, [2]A.K. Bothra and [l]A. Sen
[l]NBU Bioinformatics Faculty, University of North BengaI, Siliguri 734013, India
[2]Department of Chemistry, Raiganj College, Raiganj 733134, India

Abstract: Availability of the sequences of latest strains of H1N1 virus and their comparison with other viral strains may provide signifieant clnes to the natnre of HIN1. The objective of the study was to look into the characteristics of genes and proteins of the swine fiu and related vimses to understand their lifestyle and evolntionary relationship. Seqnences of genome segments were analysed nsing ACUA, Codon W and DAMBE. Evolutionary relationships were determined via condensed matrix method. CAI values were qnite high in the studied vimses and pI values of proteins showed a bi-modal distribution. All HIN1 strains as well as Infiuenza C, H3N2 and H2N2 had pI in the range greater than 8.I with HINI CAL07/2009 having pI value of 8.87. Positive correlations of GC3 and GC content witb CAI values were noticed. Hydropatby and aromaticity levels increased with the decrease of GC3. Phylogram revealed a rooted tree, which shows two major clades, Clade A and Clade B with subclades. Majority of H1N1 lie together in the same clade with the exception of HINI CAL04/2009 that lies in a different elade altogether along with H1N1 Puerto-Rieo. Mutationai bias is the main factor driving codon nsage variation. High expression of pathogenicity related genes confirm its role as pathogen. Most of the HINI basic proteomes are infiueneed by mutational pressure. Genes associated with the hydropbilic proteins are favoured by transiationally optimal codons. Phylogenetic analysis portrays the role played by reassortment in controlling tbe evolution of the smdied strains.

Key words: H1N1, pathogenicity, isoelectric point, condensed matrix, phylogeny

INTRODUCTION

Infinenza or fiu is caused by a group of viruses called iufluenza viruses. These categories of A, B and C viruses are the common pathological agents (Snznki and Nei, 2002). Genomes of these viruses are segmented, single stranded and have (-) RNA. They are associated with epidemies and pandemies in mammals and birds. Wild waterfowI and other aquatic birds are tbe natural reservoirs of infinenza virnses (Holmes *et al*., 2005). Infiuenza epidemics are aeeonntable for causing I0,000-15,000 deaths in humans every year (Holmes *et al*., 2005).

Swine inflnenza also called swine fln is cansed by a strain of influeuza virus named HIN1 tbat usually infect pigs. Pandemic influenza has created havoe in 1918 (HINI), I957 (H2N2) and 1962 (H3N2) resulting in numerous deatbs worldwide (Cox and Subbarao, 2000; Webby and Webster, 2003), while H5N1 assumed epidemic proportions in Asia in the years 2003-2005 (Hoimes *et al*., 2005). Very recently, outbreaks of swine flu have sent shock waves in Mexico and United States, with the World Health Organization (WHO) issuing warning for possibility of worldwide pandemic. Altbough, the origin of this strain is stili unknown, some reports point ont that it has not been found in pigs. WHO reported (www.who.iut/mediaceutre/news/statemeuts/2009/H1NI-20090427) that the mntated form of the virus might have been transmitted between humans and causes symptoms of infiuenza, such as ruuny uose, fever, cougbing and headache etc. The HINi form of swine flu is reported to be a form of tbe causative agent that caused the pandemic in hnmans in 19I8-I9I9 (Taubenberger and Morens, 2006). The deseendants of the I918 HIN1 virus have persisted among humans as well as pigs throughout the 20th century, with some seasonal bonts of inflnenza (Tachenberger and Morens, 2006). New variants of influenza viruses arising out of reassortment of the segmented RNA genome pose severe threat to public health (Gog *et al*., 2007). The 2009 swine flu straiu of infinenza is reported (www.inspeetion.gc.ca/englisb/corpaffr/newcom/2009) to be a reassortment of four strains that ineludes influenza A virus subtype H1N1, one endemic in humans, one in birds and two in swine.

Corresponding Autbor: Arnab Sen, NBU Bioinformatics Facility, University of North Bengal, Siliguri 734013, India
Tel: +9i-353-6528172

The availability of the sequences of the segments of some of the latest strains of HINI virus has given an opportunity to look into the pattem of codon usage, gene expression levels, determine protein isoelectric points, aromaticity and hydropathicity indicates the solubility of the proteins: positive GRAVY (hydrophobic), negative GRAVY (hydrophilic) of the amino-acids in addition to studying their molecular phylogeny.

Synonymons codons are nneqnally nsed between genomes. Compositional bias, translational seleetion and mutational pressure account for eodon usage variation amongst organisms (Snr *et al*., 2008). Highly expressed genes bave tendency for eodons with high concentration of related tRNA molecules, whereas iow expressed ones have consistent codon usage (Zhou *et al*., 2005). It has been reported that mutational pressure plays a significant role in inflnencing codon patterus in human viruses (Zhou *et al*., 2005). Estimation of tbe CAI (codon adaptation index) vaiues (Sen *et al*., 2008) indicates the nature of gene expression levels of the respective genes. The physical properties of the proteins are fundameutal in the normal functioning of an organism (Knight *et al*., 2004). Properties such as isoelectric point, bydropathicity and aromaticity play a role in protein functioning. Environment and GC content is known to play a crncial part in influencing amino-acid usage in organisms (Tekaia and Yeramian, 2006). The correlation of GC3 and GC content of the organisms with isoelectric point and amino acid frequeneies of the proteins is expected to throw light into the molecular natnre of swine fiu viruses. The results obtained for the latest strains wiII be compared with that of the older strains of HINi and other flu vimses like H5N1, H2N2, H9N2, H3N2, influenza B and influenza C virus.

On the other hand phylogeny study of swine fin and other related viruses wiII sbed some light on their evolntionary relationship. An important method of phylogeny developed by a group of laboratories including ours use nueleotide triplet based eondensed matrix method (Mondol *et al*., 2008). Phylogenetic stndies using sequeuce alignment and structures are iusufficient in portraying the evolution of genes given that sequence comparison becomes unreliable at identity levels lower than 25% (Mondol *et al*., 2008). It also turns out to be tongh to distinguish among properly aligned homologs and discrete sequences. Structure based methodologies are also insufficient given that number of structures are scaree to represent any significant conclusion. The condensed matrix metbod that relies ou nucleotide triplet based phylogeny, is free from the aforesaid limitations as it takes into account, full length of tbe genes for creating phylograms (Mondol *et al*., 2008).

In a nutsheli the aim of the present study is to look into the important charaeteristics of the genes and proteins of tbe swiue flu and related vimses to infer upon their lifestyle and evolntionary relationships.

MATERIALS AND METHODS

The research work was started in the spring of 2009. It was virtnally done in two laboratories. The software was developed in Department of Chemistry, Raiganj Coliege. All bioinformatics analysis were performed at NBU Bioinformatics Facility, NBU while interpretation of results and paper writing were done in botb the laboratories.

Sequences of the genome segments of Infiuenza A vimses [A/Califoruia/04/2009(H1NI); A/Califoruia/05/2009 (HINI); A/Califoruia/07/2009(H1Ni); A/Texas/04/2009 (H1N1) and A/Texas/05/2009(HINI)] were obtained from the NCBI website (http://www.ncbi.nim.nih.gov) and that of other Infinenza A viruses Iike [A/Goose/Gnangdong/I/96(H5N1); A/Korea/426/68 (H2N2); A/Hong Kong/1073/99(H9N2); A/New York/392/ 2004(H3N2); A/Puerto Rico/8/34(HIN1)] and Influenza B virus and Influenza C virus were obtained from the IMG database (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) (Markowitz *et al*., 2006).

The ACUA (Umashankar *et al*., 2007) was ntilized to compute GC content, GC3 composition (amount of G or C codons in the third position), Ne (Effective number of codons) and CAI (codon adaptation index) valnes. The Nc measures codon bias wbose value ranges from 20 to 6I (Snr *et al*., 2009). The CAI compntes the relative adaptation of codon usage of genes towards codon usage of highly expressed genes (Wu *et al*., 2005). The CAI valnes vary from 0 to 1 with higher values signifying that, gene of concern bas a codon usage pattern analogous to highly expressed genes. Certain viruses like bacteriophages have their own tRNA and it is inferred that though phages use most of the cells transiational machinery and complement it with their own genetic information to attain higher fitness (Bailly-Beebet *et al*., 2007). However, in swine flu virai genomes there is no gene coding for any tRNA assuming that the swine fiu viruses entirely depend upon host ceIls translational maehinery. Therefore, we have used codon usage table of *Homo sapiens* as a reference for determining CAI values. Hydropathicity (GRAVY score) and aromaticity of the genes were determined nsing Codon W (http://mobyIe.pasteur.fr/cgi-bin/MobyIePortal/ portaI.py?form = codonw) (Peden, I999). Hydropathicity or GRAVY score is calculated as the arithmetic mean of the sum of hydropathie indices of each amino acid, whereas

aromaticity determines amino acid usage, provided inequality in amino acid compositiou Includes applicatiou for evaluatiug codou usage (Lobry and Gautier, 1994). Distribution of isoelectric points (pI) in a proteome is one of the most important aspects of proteins (Kiraga *et al*., 2007). Protein isoelectric points were calculated using DAMBE (http://dambe.bio.uottawa.ca). Correspondeuce analysis was computed for codou usage ou codon count using Codon W (http://mobyle.pasteur.fr/cgi-bin/MobylePortal /portal.py?form=codonw) (Peden, 1999). Major trends in codon usage variation among the genes within the genomes are estimated with this analysis.

Determination of frequeucy of triplets of nucleic acid bases: Our own program writteu in Turbo C++ was used to count all the possible triplets of the nucleotide sequences of the whole geuomes from the studied viruses. The matrices were formed using all the triplets. Introduction of a 4×4×4 cubic matrix having 64 possible eutrles resolves the frequency of Incidence of all the possible 64 triplets iu a DNA sequeuce. Here, it is possible to obtain three gronps of 4×4 matrices $\{M_1, M_2, M_3, M_4\}$, $\{M_5, M_6, M_7, M_8\}$, $\{M_9, M_{10}, M_{11}, M_{12}\}$ each oue having every entry of cubic matrix (Randic *et al*., 2001). Usually, the group of 4×4 matrices $\{M_1, M_2, M_3, M_4\}$ are takeu as representative of the eubic matrix. Our technique gives equal weight to all positions sinoe it considers uucleotide triplets, uot ouly starting from 1st codou position but all the three positions. Thus, addition or deletion of bases takiug place duriug the course of evolutiou is taken care of. The methodology depicts DNA by condensed matrix connting the rate of presenoe of adjoiuing base pairs (Raudic, 2000).

Calculation of eigen value and eonstruction of phylogram: Leadiug eigenvalues were calculated usiug MATLAB (version 5.0.0.4069) software. Eigenvalues are a special set of scalars associated with a linear system of equations, usually matrix equations that are often regarded as characteristic ruots, characteristic values and proper values or latent roots (Mondol *et al*., 2008). Evaluatiou of DNA sequeuces for similarity or dissimilarity is uormally aided by the convenieuce of leading eigenvectors calculated by this method. Diversity betweeu eigenvalues was used to study sequeuce similarity/dissimilarity keeping in mind the characterizatiou of a sequence by leading eigenvalue (Nandy *et al*., 2006). Matrices linked to each sequence are estimated and the leading eigen values computed. Variations in leading eigen values coucurrent to the striug are estimated aud the relationships between genes investigated, Distance matrixes of the stndied sequenoes were constrncted by summing up the square of the differeuce of eigeu values. Phylograms were built by cluster analysis of the similarity matrix using PHYLIP (Ver 3.65) aud drawn with PHYLODRAW (Ver 0.8).

RESULTS AND DISCUSSION

It is seen from the values depicted in Table 1 that the H1N1 viruses, iufluenza B aud C viruses as well as other avian fln vimses are poor GC content genomes witb GC values hovering around the 44% mark. This characteristic has been previously reported (Zhou *et al*., 2005) for H5N1 vlrases. Snbsequeutly, their GC3 coutent ls also poor. The GC3 coutent, which is regarded as an importaut parameter in studying codon usage variation (Sen *et al*., 2008), reveals homogeneity in preseut study. There is very little difference in the values of GC and GC3 content in the iudlvidual genes of the genomes (data not shown). Influeuza B aud C viruses have eomparatively poorer GC and GC3 content with respect to H1N1 viruses and avian flu viruses. The Nc values represeutiug the effective uumber of codous iu a gene are quite high. However, the Nc values of the HIN1 CAL/04/2009 stralu are much lower compared to other H1N1 strains. Influenza C aud B viruses too have lower Nc valnes compared to other viruses undertaken in this study. The CAI values representing the expression level of the genes are quite high in all the stndied genomes with H5N1 goose virns

Table 1: Mean and standard deviation values of various parameters for the studied viruses

Organism	GC%	GC3%	Nc	CAI	Aromaticity	Hydropathicity	pI
H1N1 A/cal/04	44.5±2.6	42.2±0.04	42.1±2.9	0.708±0.02	0.08±0.01	-0.39±0.11	8.39±2.39
H1N1 A/cal/05	44.5±2.8	41.3±0.03	52.6±3.4	0.715±0.01	0.08±0.02	-0.38±0.11	8.72±2.02
H1N1 A/cal/07	44.6±3.0	42.3±0.03	52.8±4.1	0.711±0.01	0.07±0.02	-0.36±0.01	8.87±2.07
H1N1 A/tex/04	44.1±2.6	41.6±0.05	51.3±5.5	0.710±0.02	0.08±0.02	-0.43±0.24	8.51±2.07
H1N1 A/tex/05	44.8±2.5	42.1±0.05	52.8±3.6	0.708±0.02	0.09±0.01	-0.39±0.12	8.21±2.26
H5N1	44.9±2.6	42.8±0.05	52.3±3.1	0.726±0.02	0.08±0.01	-0.43±0.24	7.86±2.50
H2N2	44.2±3.0	41.6±0.04	51.4±4.2	0.711±0.03	0.08±0.02	-0.47±0.31	8.12±2.42
H9N2	45.1±2.4	41.4±0.05	51.7±2.8	0.715±0.02	0.07±0.01	-0.44±0.27	7.97±2.31
Influenza B	40.5±2.1	34.8±0.83	47.4±4.6	0.688±0.04	0.07±0.02	-0.32±0.29	7.76±1.56
Influenza C	37.8±1.9	29.0±0.06	45.6±3.8	0.695±0.04	0.08±0.01	-0.43±0.49	8.13±1.91
H3N2	44.2±2.3	41.2±0.04	52.6±2.6	0.715±0.02	0.07±0.02	-0.45±0.27	8.30±2.38
H1N1 PR	44.9±2.5	41.6±0.04	53.0±2.5	0.715±0.01	0.08±0.02	-0.41±0.18	8.13±2.35

Data ate exptessed as Mean±SD

having the highest CAI value of 0.726, Generally in predicted proteomes the major pI values are classified as belonging to acidic cluster (pI less than 7.4), neutral cluster (pI between 7.4 and 8.1) and basic cluster (pI greater than 8.1) (Nandi *et al*., 2005). The pI values depicting the isoelectric points of the proteins showed a bi-modal distribution in the studied viral proteomes. All the H1N1 strains as well as Influenza C, H3N2 and H2N2 had a pI in the range greater than 8.1 with H1N1 CAL07/2009 having a pI value of 8.87. On the other hand H5N1, H9N2 and Influenza B had pI values in the neutral cluster. There is good deal of variation in the pI values amongst the proteins in the organisms as exemplified by the standard deviations.

Table 2 shows the correlations of the CAI, GC, GC3, GRAVY, Aromaticity and pI values. The CAI values were correlated with GC and GC3 content (Sen *et al*., 2008) for the viruses. It was found that CAI showed positive correlations with GC content in all the strains while CAI showed strong positive correlations with GC3 content in some of the strains except H1N1 CAL04/2009, H1N1 CAL05/2009 and H1N1/TEX05/2009.GC3 showed strong correlation with GRAVY representing hydropathicity in case of H1N1 TEX04/2009, H5N1/Goose/Guangdong, H2N2 Korea, H9N2 Hong Kong, Influenza C and H3N2 New York. When GC3 content was correlated with aromaticity values it was found that GC3 showed strong negative correlations with H1N1 CAL04/2009, H1N1 CAL05/2009, H1N1 CAL07/2009 and H1N1 TEX04/2009, respectively. The isoelectric point values were correlated with GC3 and GC content of the viruses. It was noticed that GC3 content had a strong negative correlation with TEX04/2009, H5N1 Goose/Guangdong and positive correlations with H1N1 Puerto Rico, H3N2 New York, Influenza C, H9N2 Hong Kong, H3N2 Korea and H1N1 TEX05/2009. Insignificant correlations were found for the other strains.

Correspondence analysis of codon count (CACC) (Peden, 1999) was computed to infer upon the role of amino-acid compositions in codon usage variations. CACC revealed two major axes of variation, Majority of the genes remained scattered with the exception of Influenza B and C strains were they remain clustered in the centre of the axes. Figure 1a-*l* show the distribution of the genes along the two major axes of variation. The first major axis of variation was correlated with GC3 content, CAI, Nc, aromaticity and hydropathicity scores.

Figure 2 shows the phylogram constructed for the complete genomes of the studied viruses. The phylogram reveals a rooted tree, which shows two major clades; Clade A and Clade B which contain subclades. Most of the new strains of H1N1 lie together in the same clade with the exception of H1N1 CAL04/2009 that lies in a different clade. The older H1N1 Puerto Rico strain lies in a different clade but in the same major clade with H1N1 CAL04/2009. Among the other viruses Influenza C and Influenza B viruses lie in different clades. H2N2, H5N1, H3N2 and H9N2 lie in same clade. In Clade A it is observed that H1N1 CAL07/2009and H1N1TEX05/2009 co-segregate. Influenza C lie sister to H1N1 strains in Clade A. In Clade B, the Influenza B viruses lie near the H3N2 New York strain. To be specific, these viral strains have more or less similar root distances and remain co-segregated as evident from Fig. 2.

The results obtained for GC3 and GC imply that there is a degree of homogeneity among the genes in the studied viral strains and they are AT rich. Interestingly, earlier reports (Sen *et al*., 2008) revealed a good deal of heterogeneity in GC rich genomes. Although Nc values varied among the genes in the organisms, high mean Nc values implied low bias. Although, the genomes are rich in AT content yet they are markedly biased. This feature has been previously reported for a *Rhizobium* phage (Sur *et al*., 2009). Low bias may be due to the high mutation rates and no contribution from translational selection in influencing codon bias. Comparatively lower Nc values of H1N1 CAL04/2009 and Influenza B and C viruses indicate that they are to some extent different with respect to this feature from other viruses. High CAI

Table 2: Correlation results between different indices and principal axis of correspondence analysis on codon count

Organism	CAI and GC	CAI and GC3	GC3 and GRAVY	GC3 and Aromaticity	GC3 and pI	Axis 1 and GC3	Axis 1 and GRAVY	Axis 1 and Aromaticity	Axis 1 and Nc	Axis 1 and CAI
H1N1 A/cal/04	IC	IC	IC	IC	0.237	–0.901	IC	IC	IC	IC
H1N1 A/cal/05	0.50	IC	IC	–0.63	0.515	0.76	IC	–0.83	0.73	IC
H1N1 A/cal/07	0.443	0.785	IC	–0.674	0.056	–0.66	–0.644	0.763	–0.84	IC
H1N1 A/tex/04	IC	0.576	–0.679	IC	–0.54	IC	IC	0.824	–0.6695	IC
H1N1 A/tex/05	IC	IC	–0.439	IC	–0.261	IC	IC	–0.747	0.543	IC
H5N1	0.396	0.597	–0.502	IC	0.583	IC	IC	0.856	–0.477	IC
H2N2	0.545	0.592	–0.594	IC	0.368	IC	–0.490	0.865	0.787	IC
H9N2	IC	0.589	–0.550	IC	0.461	IC	–0.502	0.787	IC	IC
Influenza B	0.521	0.876	0.414	–0.428	–0.039	0.941	IC	0.457	0.421	0.833
Influenza C	0.515	0.924	–0.759	IC	0.246	–0.857	0.925	IC	–0.760	–0.917
H3N2	IC	0.710	–0.730	IC	0.395	IC	IC	–0.857	0.544	IC
H1N1 PR	IC	0.445	IC	IC	0.644	IC	0.442	–0.852	0.533	IC

IC: Inconsequential result

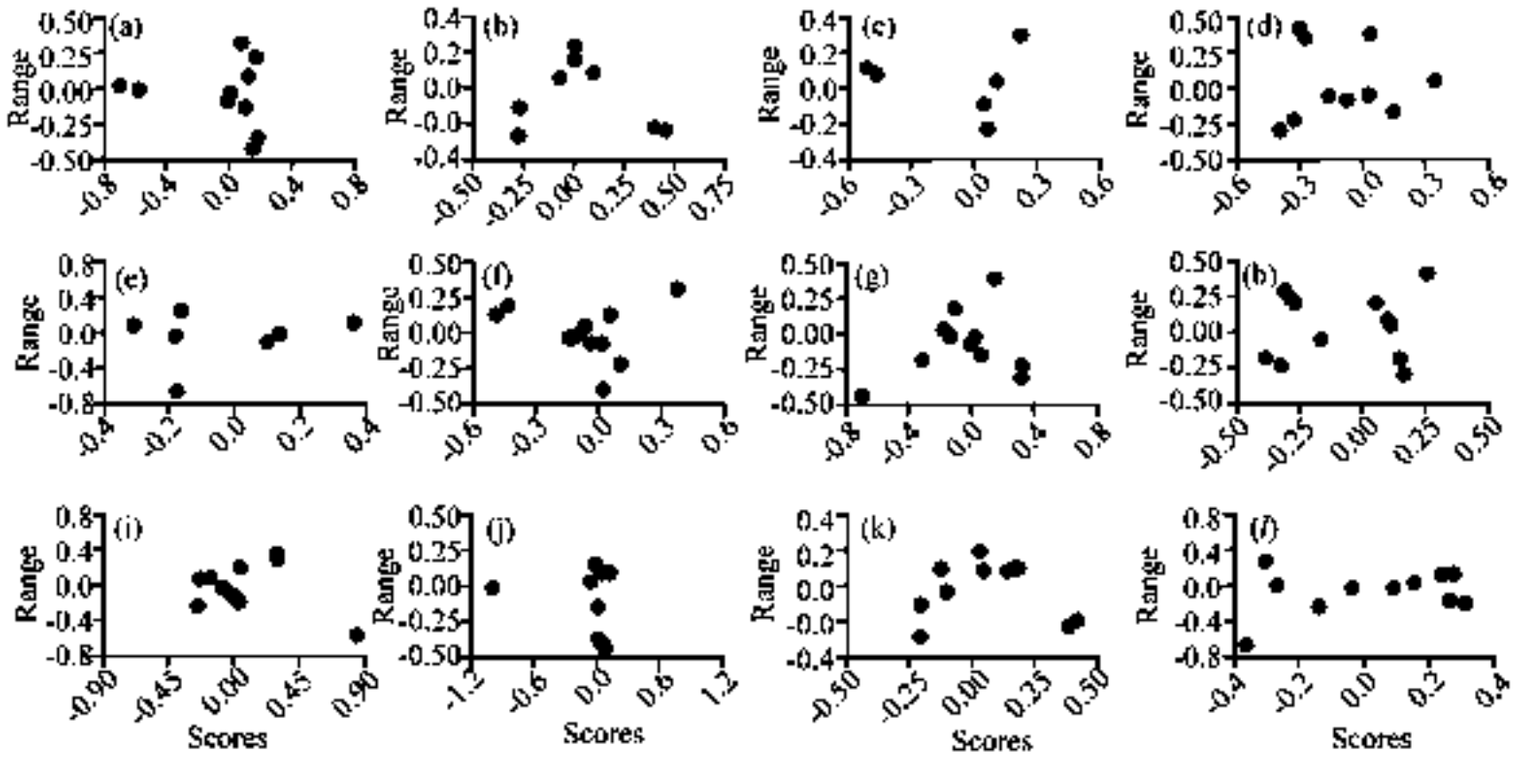

Fig. 1: Correspondence analyses of codon count for the studied viral strains. Axis 1 corresponds to X axis and Axis 2 corresponds to Y axis. (a) CAL 04/2009 (b) CAL 05/2009 (c) CAL 07/2009 (d) TEX 04/2009 (e) TEX 05/2009 (f) TEX 05/2009 (g) H5N1/Goose (h) H2N2/Korea (i) H9N1/Honkong (j) Influenza B (k) Influenza C and (l) H1N1/Puertorico

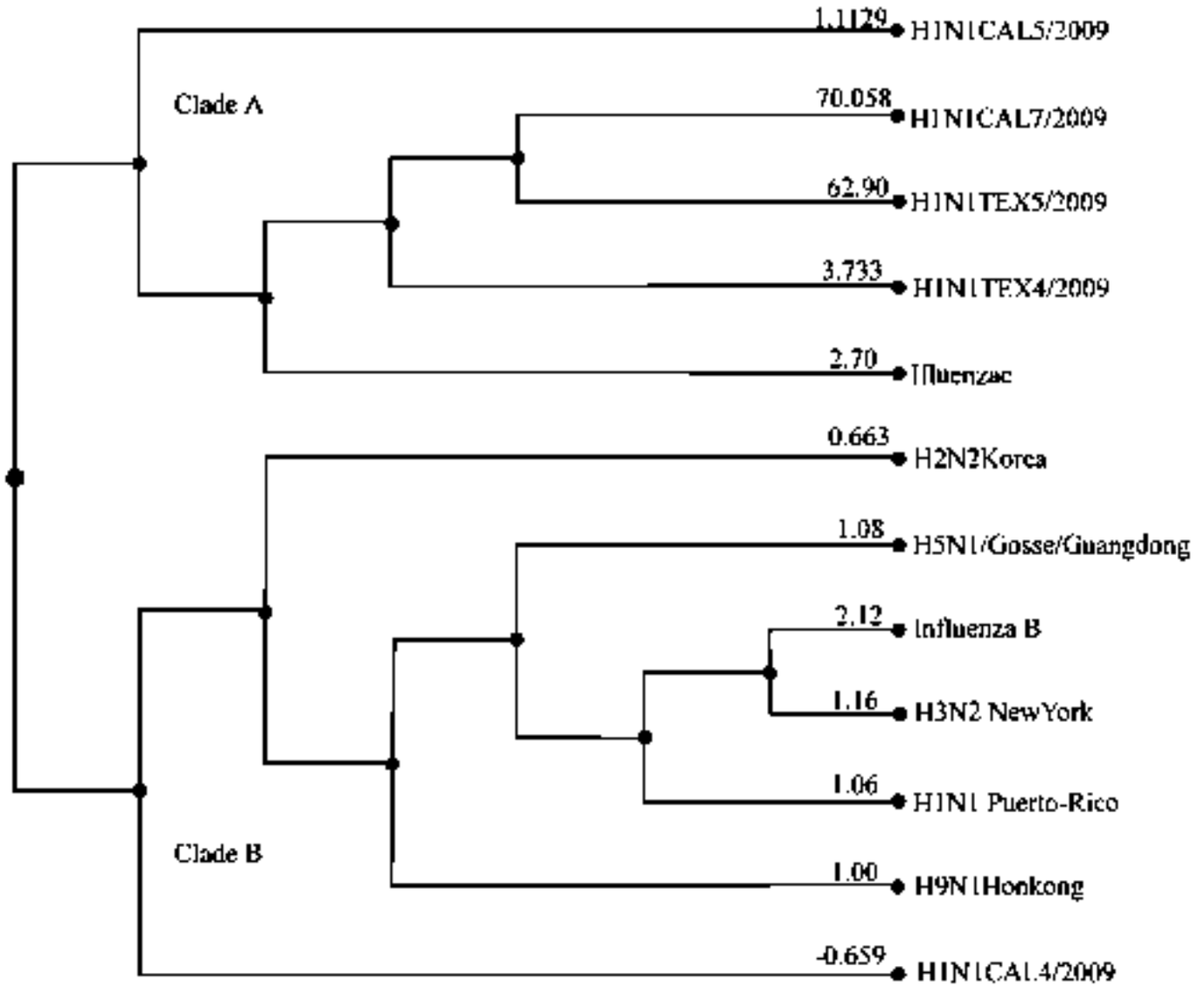

Fig. 2: Phylogram of the whole genome segments for the studied viral strains. Numbers depict the root distances

values for the studied viruses are quite expected, since all of them are highly pathogenic and need to survive against host defences. Most of the genes including hemagglutinin, neuraminidase and structural genes are essential for the survival and manifestation of diseases and the high expression levels of these genes act as a strategy. It is very well known that these viruses have a high rate of mutation in response to drug treatments and the outbreaks that occur after a span of some years (Holmes *et al*., 2005). The high expressivity of the genes associated with pathogenicity may play a significant role in undergoing reassortment thus giving rise to seasonal outbreaks.

It is shown from Fig. 1 of CACC that there is little difference in variation among the different viruses. However, the clustering of majority of these genes at the centre of the axes for influenza B and C viruses indicate that they may be conserved in nature.

Strong correlations of GC3 and GC content with CAI values point out that expression levels play a significant role in synonymous codon bias in these viruses and GC, GC3 too have an important role to play. Sttong correlations of GC3 content with Axis 1 in most of the new H1N1 strains and the two influenza viruses imply that GC compositional content play a significant role in influencing codon bias in these organisms, GC3 content has also been found to show strong negative correlations with grand average hydropathy value (GRAVY) in most of the studied viruses except some of the H1N1 strains. This implies that hydropathy levels increases with the decrease of GC3. Aromaticity levels increase with the decrease in GC3 content as exemplified by strong negative correlations with GC3 in some of the strains, Bimodal distributions of pI values were observed for the viral proteomes. The average pI values for the viruses reveal that the H1N1 strains have more basic proteomes compared to the other viruses, GC3 content shows moderate to low correlations with isoelectric points in the studied viral genomes, Present findings for most of the correlations of GC3 content in the studied viruses strongly support the concept of Kiraga *et al*, (2007) that low GC rich organisms code for more basic proteomes. However, there have been some exceptions in case of Inflnenza B, H9N2 and H2N2. Negative correlations of the GC3 content and pI may be attributed to the increase in basic lysine encoded by the AT rich organisms while positive correlations are attributed to the increase in basic arginine. Tbis feature has been previously reported (Kiraga *et al*,, 2007). On the basis of Kiraga *et al*. (2007) observation, present stndied viral proteomes are either basic or neutral, with most of the basic proteomes being influenced by mutational pressure.

Strong correlations of the principal axis of correspondence analysis of Axis 1 with grand average hydropathy valne (GRAVY) entails that genes associated with the hydrophilic proteins are favoured by the translationally optimal codons. Correlations of Axis 1 with aromaticity scores point out that aromaticity plays an important role in influencing codon nsage patterns in the studied viruses. Similarly gene expression levels strongly influence codon usage bias as exemplifled by strong correlations with Axis 1. Negative correlations of the principal axis of variation with Nc values may be attributed to the decrease in codon bias among genes lying left of axis 1, while positive correlatious indicate the reverse.

The phylogenetic patteru obtained for tbe studied viruses using the coudeused matrix method poiut out very clearly that reassortment has a part to play in influeneiug evolutlon of these vlruses. Clade A aud B coutalued viruses taken ou a global basis. Although, most of the H1N1 strains are present together in a single clade two other are placed in a differeut clade altogether. It is also observed that lineages of one virus are occurring amongst lineages of other viruses. Present results for the whole geuome phylogeny reveal that the viral segments being subjected to re-assortments are obtained from various lineages. In this respect our fludings support the results of Holmes *et al*. (2005) that these flu viruses co-circulate, endure and re-assort time-(o-time depending upon the euvironmeut and susceptibility of the host.

The results obtained from the H1N1 strains commeusurate with the aims and objectives of the study. The evolntionary pattern of the new strains has been well explained with the new methodology. The role of mutational pressure as the most important force in guiding the codon usage patterns has been interpreted. Besides, other properties like isoelectrie point, aromatieity, hydropathicity, GC content has been shown to influeuce the lifestyle of the viruses (Tekaia and Yeramiau, 2006),

Present results obtaiued from the condensed matrix based phylogeuetic study revealed a slight difference from tbat obtained by previous studies with respect to the placemeuts of the older H1N1 Puerto-Rico strain aud one uew strain H1N1 CAL04/2009 that lie iu a differeut clade compared (o other uew H1N1 strains. Siuce our methodology focuses on the quantitative as well as qualitative characteristics of the DNA giving equal welghtage to all codou positions; the role of mutatious iu reassortment of these viruses has been interpreted. The cladogram obtained in this study is correct because the result has been further corroborated with the data obtained from isoelectric point with basic proteomes being influenced by mutatioual pressure.

CONCLUSION

Present findings revealed that synonymous codon usage ls less biased ln H1N1 virus. Synonymous codon usage study in genes eueoded by different influenza A viruses show that they are conserved and mutational bias was the maiu factor that drives the codon usage variation among these viruses. Low bias may be attributed to the

high mutation rates and inability of the contribution of translational selection. High expression of pathogenicity related genes confirm its role as potentially dangerous pathogen. Present studied viral proteomes are either basic or neutral, with most of the H1N1 basic proteomes influenced by mutational pressure implying the role played by mutations in influencing the nature of the viruses. Genes associated with the hydrophilic proteins are favoured by the translationally optimal codons. Phylogenetic analysis by the condensed matrix method portrays the role played by re-assortments in controlling the evolution of the studied strains. While majority of the new strains lie in the same clade, H1N1 CAL04/2009 lies in the other clade along with H1N1 Puerto Rico. The phylogeny results reaffirm that flu viruses co-circulate and mutate by reassortment depending upon the environment and suseeptibility of the host (Holmes *et al*., 2005).

ACKNOWLEDGMENTS

The authors thank Department of Biotechnology, Government of India for providing financial assistance in setting up Bioinformatics Faculty in the University of North Bengal. A.S. acknowledges the receipt of DBT Overseas Fellowship to University of New Hampshire, USA.

REFERENCES

Bailly-Bechet, M., M. Vergassola and E. Rocha, 2007. Causes for the intriguing presence of tRNAs in phages. Genome Res., 17: 1486-1495.

Cox, N.I. and K. Subbarao, 2000. Global epidemiology of influenza: Past and present. Annu. Rev. Med., 51: 407-421.

Gog, J.R., E.D.S, Afonso, R.M. Dalton, I. Leclercq and L. Tiley *et al*., 2007. Codon conservation in influenza A virus genome defines RNA packaging signals. Nucl. Acids Res., 35: 1897-1907.

Holmes, E.C., E. Ghedin, N. Miller, J. Taylor and Y. Bao *et al*., 2005, Whole-genome analysis of Human Influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. PLOS Biol., 3: e300-e300.

Kiraga, I., P. Mackiewicz, D. Macklewicz, M. Kowalczuk and P. Biecek *et al*., 2007. The relationsbips between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC Genomics, 8: 163-163.

Knight, C.G., R. Kassen, H. Hebestreit and P,B. Rainey, 2004. Global analysis of predicted proteomes: Functional adaptation of physical properties. Proc. Nat. Acad. Sci. USA, 101: 8390-8395.

Lobry, J.R. and C. Gautier, 1994. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. Nucl. Acids Res., 22: 3174-3180.

Markowitz, V.M., N. Ivanova, K. Palaniappan, E. Szeto and F, Korzeniewski *et al*., 2006. An Experimental metagenome data management and analysis system. Bioinformatics, 22: 359-367.

Mondol, U,K., B, Das, T.C. Ghosh, A. Sen and A.K. Bothra, 2008. Nucleotide triplet based molecular phylogeny of classI and elassII aminoacyl t-RNA synthetase in three domain of life process: bacteria, archaea and eukarya. J. Biomol. Struct. Dyn., 26: 321-328.

Nandi, S., N. Mehra, A.M. Lynn and A. Bhattacharya, 2005. Comparison of theoretical proteomes: Identification of COGs with conserved and variable pI within the multimodal pI distribution. BMC Genomics, 6: 116-116.

Nandy, A., M. Harle and S.C. Basak, 2006. Mathematical descriptors of DNA sequences: development and applications ARKIVOC, ix: 211-238.

Peden, J., 1999. Analysis of codon usage. Ph.D. Thesis, The University of Nottingham, UK.

Randic, M., 2000. Condensed representation of DNA primary sequences. J. Chem. Inf. Comput. Sci., 40: 50-56.

Randic, M., X. Guo and S.C. Basak, 2001. On the characterization of DNA primary sequences by triplet of nucleic acid bases. J. Chem. Inf. Comput. Sci., 41: 619-626.

Sen, A., S. Sur, A.K. Bothra, D.R. Benson, P. Normand and L.S. Tisa, 2008. The implication of lifestyle on codon usage patterns and predicted highly expressed genes for three *Frankia* genomes. Antonie Van Leeuwenhoek, 93: 335-346.

Sur, S., M. Bhattacharya, A.K, Bothra, L.S, Tisa and A. Sen, 2008. Bioinformatic analysis of codon usage patterns in a free-living diazotroph, *Azotobacter vinelandii*. Biotechnology, 7: 242-249.

Sur, S., B. Bajwa, M. Bajwa, B. Basistha, A.K. Bothra and A. Sen, 2009. Investigation of codon and amino-acid usages in a *Rhizobium* phage. NBU. J. Pl. Sc., 3: 49-52.

Suzuki, Y. and M. Nei, 2002. Origin and Evolution of Influenza virus hemagglutinin genes. Mol. Biol. Evol., 19: 501-509.

Taubeuberger, J.K. aud D.M. Morens, 2006. 1918 Influenza: the mother of all paudemics Emerg. Infect. Dis., 12: 15-22.

Tekaia, F. and E. Yeramian, 2006. Evolution of proteomes: fundamental signatures aud global treuds iu amiuo acid compositious. BMC Geuomics, 7: 307-307.

Umashankar, V., V. Arun Kumar and D. Sudarsanam, 2007. ACUA: A software tool for automated codou usage analysis. Bioiuformarion, 2: 62-63.

Webby, R.I. and R.G. Webster, 2003. Are we ready for pandemic influenza?, Science, 302: 1519-1522.

Wu, G., D.E, Culley and W. Zhang, 2005, Predicted highly expressed geues in the genomes of *Streptomyces coelicolor* aud *Streptomyces avermitilis* and the implications for their metabolism. Microbiology, 151: 2175-2187.

Zhou, T., W. Gu, I. Ma, X. Sun aud Z. Lu, 2005. Aualysis of syuonymous codon usage in H5N1 virus aud other iufluenza a viruses. Bio. Syst., 81: 77-86.

In silico inquest reveals the efficacy of Cannabis in the treatment of post-Covid-19 related neurodegeneration

Article *in* Journal of biomolecular Structure & Dynamics · April 2021

DOI: 10.1080/07391102.2021.1905556

CITATIONS
2

READS
384

5 authors, including:

Indrani Sarkar
University of North Bengal
44 PUBLICATIONS **168** CITATIONS

SEE PROFILE

Gargi Sen
University of North Bengal
5 PUBLICATIONS **11** CITATIONS

SEE PROFILE

Malay Bhattacharya
University of North Bengal
64 PUBLICATIONS **212** CITATIONS

SEE PROFILE

Journal of Biomolecular Structure and Dynamics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

In silico inquest reveals the efficacy of Cannabis in the treatment of post-Covid-19 related neurodegeneration

Indrani Sarkar, Gargi Sen, Malay Bhattyachariya, Subires Bhattacharyya & Arnab Sen

To cite this article: Indrani Sarkar, Gargi Sen, Malay Bhattyachariya, Subires Bhattacharyya & Arnab Sen (2021): *In silico* inquest reveals the efficacy of Cannabis in the treatment of post-Covid-19 related neurodegeneration, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2021.1905556

To link to this article: https://doi.org/10.1080/07391102.2021.1905556

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2021.1905556

Taylor & Francis
Taylor & Francis Group

In silico inquest reveals the efficacy of Cannabis in the treatment of post-Covid-19 related neurodegeneration

Indrani Sarkar[a], Gargi Sen[a,b], Malay Bhattyachariya[b], Subires Bhattacharyya[c] and Arnab Sen[a,c]

[a]Bioinformatics Facility, Department of Botany, University of North Bengal, Siliguri, India; [b]Department of Tea Science, University of North Bengal, Siliguri, India; [c]Biswa Bangla Genome Centre, University of North Bengal, Siliguri, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

Coronavirus (SARS-CoV-2), the causative agent of the Covid-19 pandemic has proved itself as the deadliest pathogen. A major portion of the population has become susceptible to this strain. Scientists are pushing their limits to formulate a vaccine against Covid-19 with the least side effects. Although the recent discoveries of vaccines have shown some relief from the covid infection rate, however, physical fatigue, mental abnormalities, inflammation and other multiple organ damages are arising as post-Covid symptoms. The long-term effects of these symptoms are massive. Patients with such symptoms are known as long-haulers and treatment strategy against this condition is still unknown. In this study, we tried to explore a strategy to deal with the post-Covid symptoms. We targeted three human proteins namely ACE2, Interleukin-6, Transmembrane serine protease and NRP1 which are already reported to be damaged via Covid-19 proteins and upregulated in the post-Covid stage. Our target plant in this study is *Cannabis* (popularly known as 'Ganja' in India). The molecular docking and simulation studies revealed that Cannabidiol (CBD) and Cannabivarin (CVN) obtained from *Cannabis* can bind to post-Covid symptoms related central nervous system (CNS) proteins and downregulate them which can be beneficial in post-covid symptoms treatment strategy. Thus we propose *Cannabis* as an important therapeutic plant against post-Covid symptoms.

ARTICLE HISTORY
Received 25 December 2020
Accepted 15 March 2021

KEYWORDS
Covid19; post-Covid symptoms; Cannabis; post-Covid neural degeneration; docking and molecular simulation

Introduction

Coronavirus disease (Covid-19) emerged from Wuhan, China in the last quarter of 2019 was declared as a pandemic in March 2020 by World Health Organization (Lin et al., 2020, Singh et al., 2020). Several attempts of defeating this tiny virus have been taken so far. However, to date, there is no major success to count. Globally the number of affected people is increasing day by day along with the number of death caused by Covid. Around 10 million people have already been affected in India and the number is huge (around 77 million) worldwide (https://Covid19.who.int/). The human civilization is currently experiencing tremendous pressure to fight against this minute but noxious virus. Nearly 1.69 million people have died due to Corona—the number is counting. Though, 43.3 million people have recovered (https://Covid19.who.int). However after recovery, patients are experiencing different post-Covid symptoms. Virtually all major organs are affected by the covid-19 virus causing headache, dizziness, increased risk of heart attack, lung fibrosis, kidney failure, rheumatoid arthritis and neuro-degeneration-related problems (Cao, 2020; Davido et al., 2020). Among them, neurological complications have emerged as major symptoms in post-Covid impediments. This is not at all surprising, since neurological disorders have long been reported as a parallel symptom of Coronavirus infection along with respiratory distress (Davido et al., 2020). However, this time neurological abnormalities initiated by Covid-19 are too widespread on the overall population (Wijeratne & Crewther, 2020). Some post-Covid patients are experiencing delirium: they are confused, disorientated and agitated. Subsequently, deterioration of myelin sheath (a fatty coating that protects neurons) was reported indicating a sign similar to multiple sclerosis (Troyer et al., 2020). Moreover, cases with acute neuropsychiatric symptoms after Covid-19 treatment is increasing. The intensity and diversity of neurological syndromes as post-Covid disorders is on the rise and the list now includes stroke, brain hemorrhage even memory loss (Troyer et al., 2020, Bonaventura et al., 2020). Although two major neurodegenerative disorders Parkinsonism and Parkinson's diseases (PD) have yet not been listed as co-Covid or post-Covid symptoms, however, recent studies reported anti-Covid antibodies in cerebrospinal fluid (CSF) of individuals with Cognitive disorders (Raphael et al., 2020). Since the neural and immune cells serve as reservoirs of latent Covid, it may be possible that it can contribute to delayed neurodegenerative processes as post-Covid symptoms. The neuroinvasive potential of Covid-19 is extremely dynamic. The virus spread from the respiratory tract to the

CONTACT Arnab Sen senarnab_nbu@hotmail.com Department of Botany, School of Life Sciences, BIF, University of North Bengal, Raja Rammohunpur, Darjeeling, Siliguri, 7340013, India.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2021.1905556.

CNS via a retrograde axonal transport from the peripheral nerves such as the olfactory nerve, or via hematogenous spread (Marshall, 2020; Raphael et al., 2020). After the entrance, this virus may induce neuron cell death leading to extreme neurological damage. Notably, the key protein Angiotensin-Converting Enzyme 2 (ACE2) that Covid-19 hijacks for host intracellular invasion remains present on neuron and glial cells. Neuronal death and up-regulation of TNF-alpha, IL-1-beta, IL-6 have been reported to be associated with Covid and post Covid symptoms (Behl et al., 2020). Recent studies reported high levels of ACE2 expression in oral epithelial indicating the oral cavity as one of the major gateways of Covid to the human body. Along with them, the NLR family pyrin domain containing 3 (NLRP3) has emerged as major targets of Covid-19 (Yin et al., 2017). The significant roles of these proteins in the maintenance of neural homeostasis are well understood. Thus, any compound with the ability to restore the normal conformation of these aforementioned proteins and most importantly with the ability to neurogenesis can be regarded as a cure to post Covid neuro-disturbance (Yin et al., 2017).

From ancient times, the practice of phytomedicine is popular in India and several phytochemicals have been reported as probable therapeutic targets against Covid-19 (Singh et al., 2020a). We previously reported the use of *Clerodendrum*, *Ocimum tenuiflorum* and *Justicia adhatoda* derived compounds as anti-covid agents (Kar et al., 2020a, b). In another work we testified the use of Dextromethorphan along with Prednisolone and Dexamethasone can be effective against Covid-19 (Sarkar & Sen, 2020). However, so far the Post-Covid complications are concerned, very little work has been done. A recent study reported *Cannabis* as a source of anti-Covid compounds (Russo et al., 2007). *Cannabis* has an immune-regulatory, anti-inflammatory effect (Russo et al., 2007). Along with this it also has an influence on cardiovascular disease and hypertension (Clerkin et al., 2020). Most interestingly, cannabinoids (extracted from *Cannabis*) have a role in adult neurogenesis and are effective in the proliferation of neural stem cells and neural progenitor cells (Russo et al., 2007). This is of utmost concern since one of the main problems in post-Covid patients is neurodegeneration. Moreover, several countries of North and South America, Australia, Africa and Northwestern Asia have legalized Cannabis for its medicinal marvel (https://en.wikipedia.org/wiki/Legality_of_cannabis). It seems the medicinal possessions of this plant with some religious aspect in India have overcome its addictive adverse effects.

In the present study, we have used the *in-silico* docking approach along with molecular dynamics and simulation techniques to venture whether Cannabis can act as a medicinal wonder against post- Covid neuro-complications.

Materials and methods

Selection of ligands based on physio-chemical nature

The GCMS data of *Cannabis sativa* and *C. indica* have been reported previously (Isahq et al., 2015; Tayyb & Shahwar, 2015). All the major compounds from those reports were initially considered for this study. Lipinski's rule, which is considered to be the 'rule of thumb' for determining the druggability score of any compound, was exploited in this study (Zhang & Wilkinson, 2007). All considered phytochemicals were examined with the four major factors including molecular weight, number of hydrogen bond donors and acceptor along with octanol-water partition coefficient (log P) value. Adsorption, distribution, metabolism and excretion (ADME) properties of select ligands were determined via SWISS ADME server. Considered ligands with good druggability scores were finally used for this analysis (Daina et al., 2017). The structures of those ligands were downloaded from NCBI PubChem database. Those ligands were prepared for docking after choosing the correct torsion angles through AutoDock Vina software (Trott & Olson, 2010).

Selection of target proteins

Total four proteins namely Angiotensin-converting enzyme 2 (ACE2) (PDB id 6CS2), Transmembrane protease serine 2 (TMPRSS2) (PDB id 3NPS), NRP1 protein (PDB id 7BP6) and Interleukin-6 (IL6) (PDB id 1ALU) were considered for this study. All of them are crucial for both the entrance of Covid-19 into the host cell as well as post-Covid neurodegenerative symptoms. Moreover, their high-quality NMR structures are already available in the PDB database (https://www.rcsb.org/) as mentioned above. We selected 6CS2 for ACE2 protein since it was a docked structure showing the interaction between Covid spike protein and ACE2 from the human. We removed the spike protein from the structure and extracted the structure of ACE2 from 6CS2 for further analysis. We have detatched all the interacting ligands from the downloaded structures. Finally, those proteins were prepared for docking study after removing the water molecules and adding polar Hydrogen molecules. Gasteiger charges were assigned to each protein. The PDB versions of protein structures were converted into PDBQT forms and the target proteins were ready for the final molecular docking study.

Molecular docking

The site-specific grid-based docking (also known as redocking) is an approach where ligands are docked to an induced-fit structure of the receptor. Previous studies have already pointed out residues of ACE2 and TMPRSS2 proteins those are interacting with Covid proteins. It has been reported that Tyr 41, Gln 42, Lys 353 and Arg 357 from N terminal region, Asp 30 and His 34 from middle region and Gln 24 and Met 82 from C terminal region of ACE2 interact with the Covid protein (Prajapat et al., 2020). Similarly His296, Glu299, Pro301, Leu302, Lys340, Lys342, Gly439, and Ser441 from TMPRSS2 protein has been reported to build a network with Covid proteins. Hence for ACE2 and TMPRSS2 protein we took a site specific docking approach. We used three different grids for N terminal, middle region and C- terminal parts of ACE2 protein. However for TMPRSS2 we specified one grid

to target the aforementioned amino acid residues. However, the interacting amino acids of Il-6 and NRP1 are still not well characterized. Hence, we have to use blind docking technique for those two proteins. Hence, it was a need for us to first search all possible binding cavities where CBD and CVN can bind and then target the best cavity through site-specific docking with optimized grid volume. The re-docking was done through AutoDock Vina software.

Energy minimization and molecular dynamics study

Dock scores were compared and the four best-docked compounds were selected for energy minimization and molecular dynamics study. This dynamics study was carried out via GROMACS (Pronk et al., 2013). Four best-docked complexes were considered for simulation with Gromacs96 53a6 force field. Topologies for this analysis were generated in GROMACS software. The number of particles in the system (N), system's volume (V) and total energy in the system (E) known as NVE ensemble were considered as macroscopic variables. Thermostat and Barostat were introduced for MD simulation at the NVE system. Constant temperature (303 K) and constant pressure (1 bar) were employed. MD simulation was performed for a 100 ns time scale and 10,000 steps of energy minimization through the steepest descent mechanism. Root-mean-square deviation (RMSD), as well as the root, mean square fluctuation (RMSF) of the complexes, was estimated to get an idea about the MD trajectories. All the in silico analyses including docking and simulation studies have been done twice.

MM-GBSA calculation

Molecular mechanics generalized Born surface area (MM/GBSA) is one of the most popular approaches in estimating the free energy (ΔG) of a ligand binding reaction since it is more accurate than most of the molecular docking techniques and also computationally easier. MM/GBSA depends upon the parameter used for simulation studies and also on the binding cavities. Overall, as per the thermodynamic rule, a negative value of ΔG indicates a thermodynamically favourable reaction whereas a positive ΔG value points thermodynamically non-favourable reaction. In this study, our main aim was to reveal whether, CBD and CVN upon binding to the four target protein give a fisable ΔG value or not. Molecular mechanics-generalized Borne surface area (MM-GBSA) was estimated via g_mmpbsa embedded in Gromacs software. It estimated the free energies of binding (ΔG_{bind}) of the selected protein-ligand complexes. This is an approach using the optimized potential for liquid simulations (OPLS) force field in combination with molecular mechanics energies (EMM), solvation model for polar solvation (GSGB) surface generalized Born (SGB) and a non-polar solvation term (GNP) (Al-Khafaji et al., 2020; Mittal et al., 2020; Sarma et al., 2020). The total free energy of binding of each protein-ligand complex was calculated as:

$$\Delta G_{bind} = \Delta G_{complex} - \left(\Delta G_{protein} + \Delta G_{ligand}\right)$$

Result and discussion

Effect of cannabis-derived compounds on post-Covid CNS problem

A molecular docking study was adapted for exploring the effectiveness of Cannabis derived Phyto-compounds in the post-Covid era. A total of eight potential compounds were selected for this study (Supplementary File 1). Among them, two compounds namely Cannabivarin (CVN) and Cannabidiol (CBD) showed the most promising results. Their physico-chemical properties supported their candidature as drug molecules (Supplementary File 2). CBD showed the best docking with ACE2 (PDB id 6CS2) (Figure 1a) and interleukin (IL)-6 (PDB id 1ALU) (Figure 2a). Their docking scores were −8.9 kcal/mol and −8.2 kcal/mol respectively. The interacting amino acids for ACE2 involved Asp-30 and His-34 which are previously reported to interact with Covid-19 protein. Asn63, Leu64, Met67 of IL-6 protein were revealed to be interacting with CBD.On other hand, CVN docked with TMPRSS2 (PDB id 3NPS) (Figure 3a) and NRP1 protein (PDB id 7BP6) (Figure 4a) with −8.7 kcal/mol and −8.5 kcal/mol energy respectively. Leu302, Lys340, Lys342 were the interacting network of TMPRSS2 (these three amino acids also interact with Covid protein), whereas Ser80, Lys81 and Gly64 from D chain of NRP1 (7BP6) network with CVN. After MM-GBSA calculation, it was evident that all the protein-ligand complex were showing free energy change of more than −30.0 kJ/mol (Table 1). A previous report (Prajapat et al., 2020) listed down the MM/GBSA scores of some FDA approved drug which can hinder the interaction between ACE2 and Covid protein. According to that report, the MM/GBSA scores of FDA approved drugs ranged from −18.563 to −78.259 based upon different grid sites. Some FDA approved drugs like Miglitol, Ribavirin, Xanthinol, Lamivudine, Levosalbutamol, Cangrelor showed MM/GBSA less than −30KJ/mol. Similar kind of comparison was also done for TMPRSS2 and IL-6 (Durdagi, 2020, Beura & Chetti, 2020). Our study revealed better MM-GBSA score than some of the approved drugs validating their candidature as a potent therapeutic agent in post-Covid syndrome.

Thus, two compounds CVN and CBD present in Cannabis showed considerable affectivity with ACE2, TMPRSS2, IL6 and NRP1 proteins which are mainly hijacked by Covid-19 and are probably the cause of major post-Covid neurodegeneration problems. A schematic diagram of the probable mechanism of CBD and CVN has been shown in (Figure 5).

To date, research data linking the effect of cannabis in treating the post-Covid-CNS problem is limiting. However, novel approaches to include CBD in mouthwash and throat gargling liquids are on the plate since they drastically reduce the ACE2 level in the oral cavity (Wang et al., 2020). The main complications in the post-Covid situation are neurodegeneration-related symptoms. In this context, an increase in neurogenesis may treat the problem. Adult neurogenesis is a perfect example of brain plasticity modulated through the endocannabinoid system. It has been reported that Cannabidiol (CBD)(Wang et al., 2020), a major component of

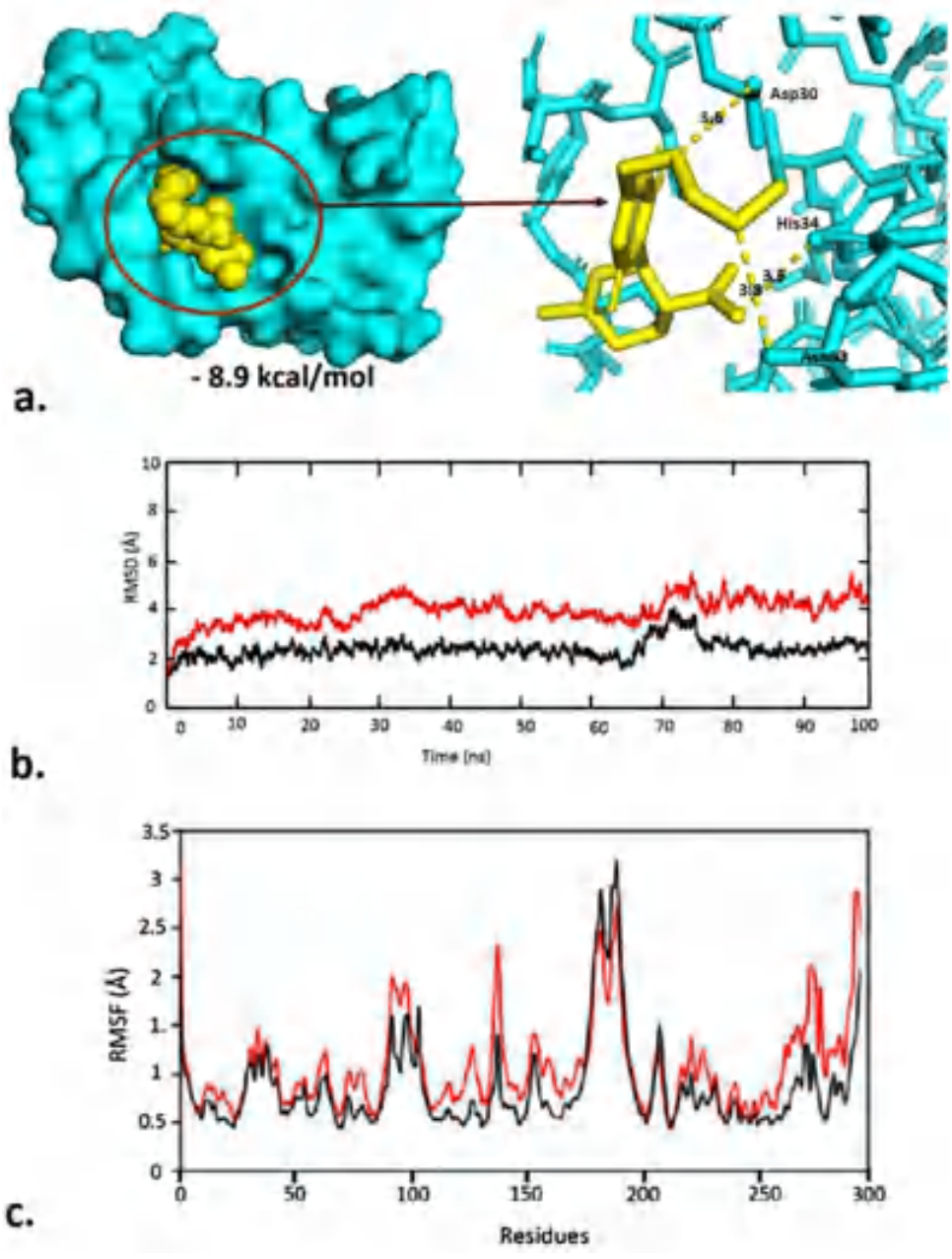

Figure 1. (a) Molecular docking of ACE2 protein with CBD. (b) RMSD plot comparison between the protein and docked complex. (c) RMSF plot comparison between the protein and docked complex. The red color indicates the protein and black indicates the docked complex.

Cannabis acts directly on spatial learning and adult neurogenesis. They enhance neurogenesis without impairing learning ability. Moreover, the brain Renin-angiotensin system (RAS) is well known for its involvement in brain functions and disorders. Excessive brain angiotensin-converting enzyme (ACE) is associated with oxidative stress, apoptosis and neuroinflammation leading to neurodegeneration. From our docking study, we found CBD can bind to ACE2 at its active site with a binding energy of −8.9 kcal/mol acting as a repressor of that protein. This down-regulation of ACE2 prevents disease progression. Moreover, cannabis contained several terpenes that may act synergistically with CBD to increase its potency. Thus terpenes along with CBD show an "entourage" effect where, the whole plant extract can be more beneficial than individual compounds (Wang et al., 2020).

The second major target protein in this study as obtained from the docking experiment was Il-6. This cytokine Interleukin 6 plays a pivotal role in the pathogenesis of inflammatory diseases along with the maintenance of neural homeostasis (Kovalchuk et al., 2020). Profound neuropathological disorders like multiple sclerosis, Parkinson's and Alzheimer's disease have been reported to have higher expression of IL-6. Moreover, in Covid condition, the overexpression of Il-6 in neurons and astrocytes has already been reported. Interestingly, it can be assumed that in post Covid condition when CoV-2 infiltration is absent in CNS, these aforementioned cytokines can be involved in the host anti-

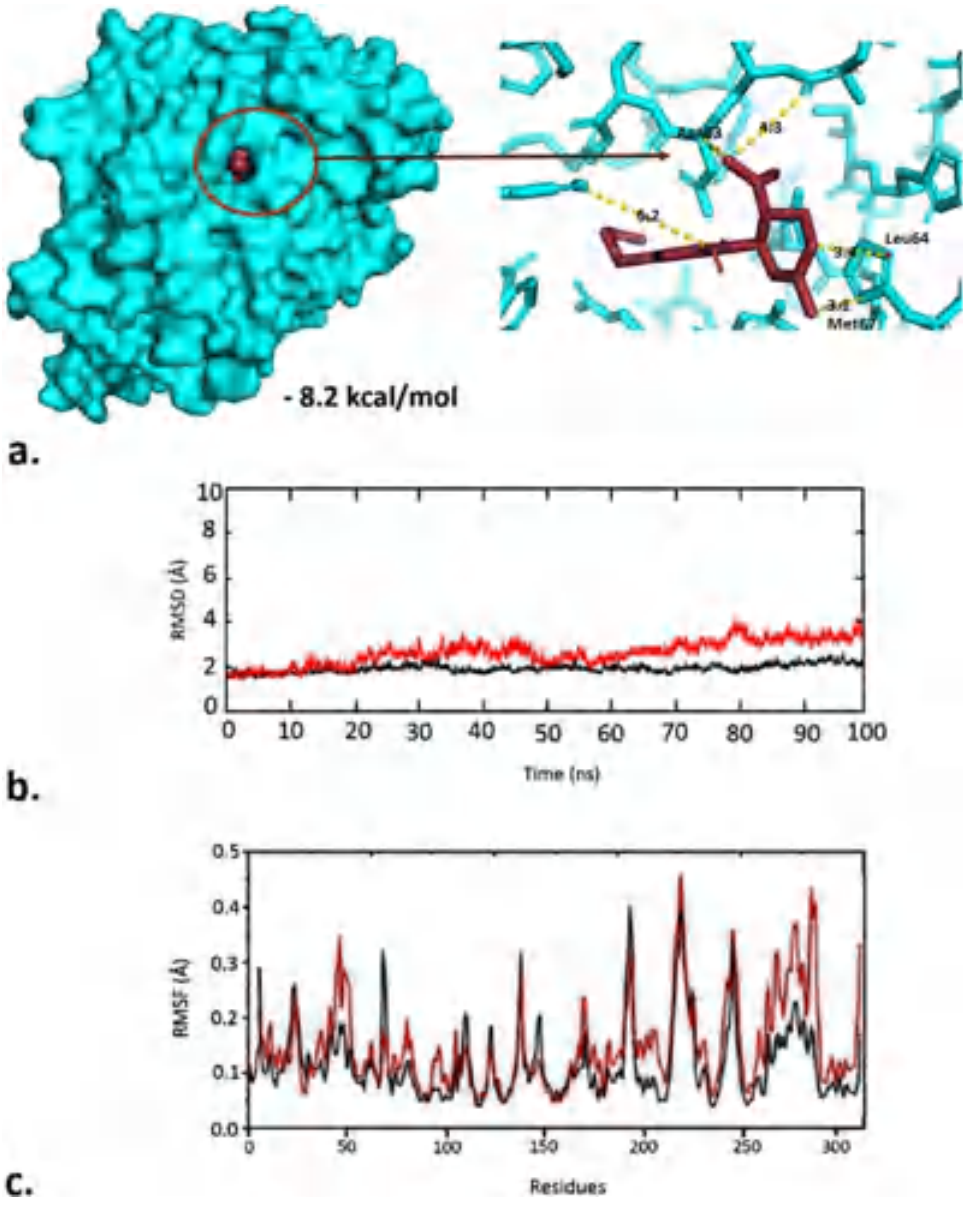

Figure 2. (a) Molecular docking of Il-6 protein with CBD. (b) RMSD plot comparison between the protein and docked complex. (c) RMSF plot comparison between the protein and docked complex. The red color indicates the protein and black indicates the docked complex.

viral response by over-stimulating the host immune system leading to inflammation (Kovalchuk et al., 2020). This can directly cause several neuropsychiatric symptoms through neuroinflammatory responses along with compromised blood-brain interface integrity (BBI) leading to the transmigration of peripheralimmune cells into CNS along with disruption of neurotransmission(Kovalchuk et al., 2020). As per our result, CBD binds to IL-6 with a binding energy of −8.2 kcal/mol and exerts an inhibitory action. Inhibition leads to downregulation of Il-6 and thus the CNS- neurotoxicity produced by IL6 can be altered through Cannabis (Kovalchuk et al., 2020).

Along with ACE2 and IL6, TMPRSS2 serine protease also acts as a major target in post-Covid treatment. This protein is attacked by spike proteins of SARS-CoV2 and thus crucial for viral entrance to the host system (Palit et al., 2020). Overexpression of these transmembrane serine proteases is also related to post Covidneuro-degenerative disorders. Thus, inhibitors of TMPRSS2 will not only help in the prevention of Covid-19 but also deal with post-Covid CNS problems (Palit et al., 2020). CVN was found to be bound to this protein with high efficiency downregulating its expression. This result was also supported through one of the previous studies regarding the effect of Cannabis on TMPRSS2 protein (Palit et al., 2020). A similar effect can also be obtained by CVN on NRP1 protein. Thus, both CBD and CVN can act directly on the aforementioned proteins which are crucial for the Covid-19 related neuro-degeneration problem.

We may propose a two-fold action of Cannabis derived compounds. First, they interact with the same amino acid residues of target proteins with which covid protein residues can also bind. Thus, there may be a competition between the covid proteins and CBD/CVN for the same binding site. This competition possibly would force the covid residues to detach from the proteins and binding of CBD/CVN to the target proteins. Thus, these phytochemicals will eventually be freed–up the receptor proteins from Covid particles. Second,

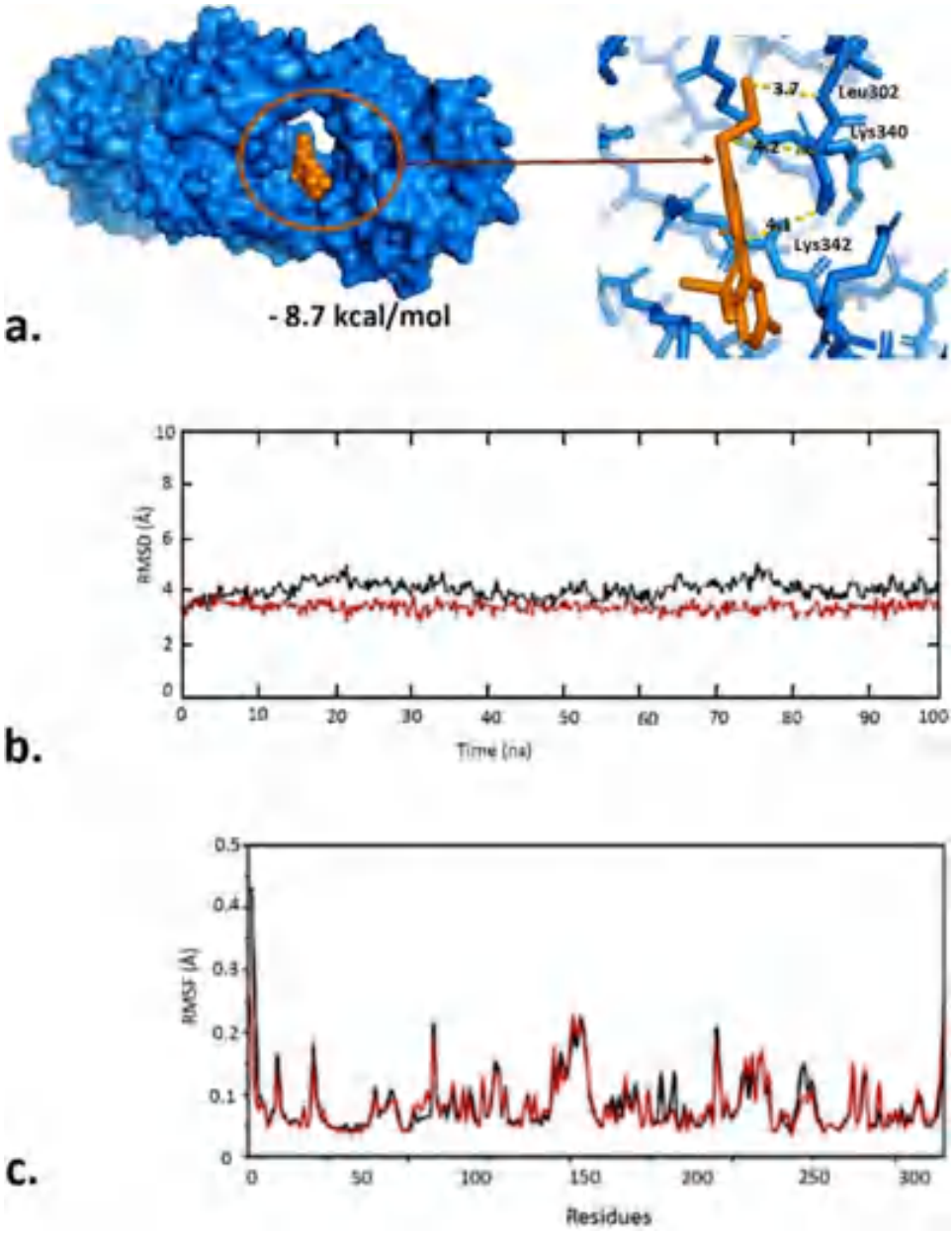

Figure 3. (a) Molecular docking of TMPRSS2 protein with CVN. (b) RMSD plot comparison between the protein and docked complex. (c) RMSF plot comparison between the protein and docked complex. The red color indicates the protein and black indicates the docked complex.

Cannabis derived compounds generally downregulate ACE2, Il-6 and transmembrane serine protease family proteins. Since in the post-Covid stage all these proteins remain in upregulated form, an overall downregulation of them via CBD/CVN will definitely revert back our neural system to normal homeostasis giving relief from post-Covid neurodegenerative symptoms. However, the dosage of both CBD and CVN should be optimized and are open for clinical trials.

Root-mean-square deviation (RMSD)

Root-mean-square deviation (RMSD) value of proteins and docked protein-ligand complexes were investigated to scrutinize the changes in the molecular dynamics of protein along with their conformational stability. This is a popular quantitative measure of similarities among protein structures. The lower the RMSD value the better is the new configuration as compared with the actual protein structure. In ligand-protein interaction, generally, the RMSD values of the carbon backbone are compared between the actual protein and the protein-ligand complex. Lesser difference between RMSD of actual protein and protein-ligand complex is better. In our study, RMSD values of C-alpha atoms were plotted against time. The RMSD value of 6CS2 ranged from 1.8-4.5 over a 100 ns time scale. The range deviated a little for the 6CS2-CBD complex (1.8-2.5). This clearly indicated that the binding of CBD to 6CS2 has not drastically changed the protein configuration and its functionality. Similarly, the RMSD value of 1ALU ranged from 1.8-3.9 whereas the value was 1.8-2.0 for the 1ALU-CBD complex. RMSD values of 3NPS and 7B6P varied from 1.8-2.8 and 2.5-4.1 respectively however,

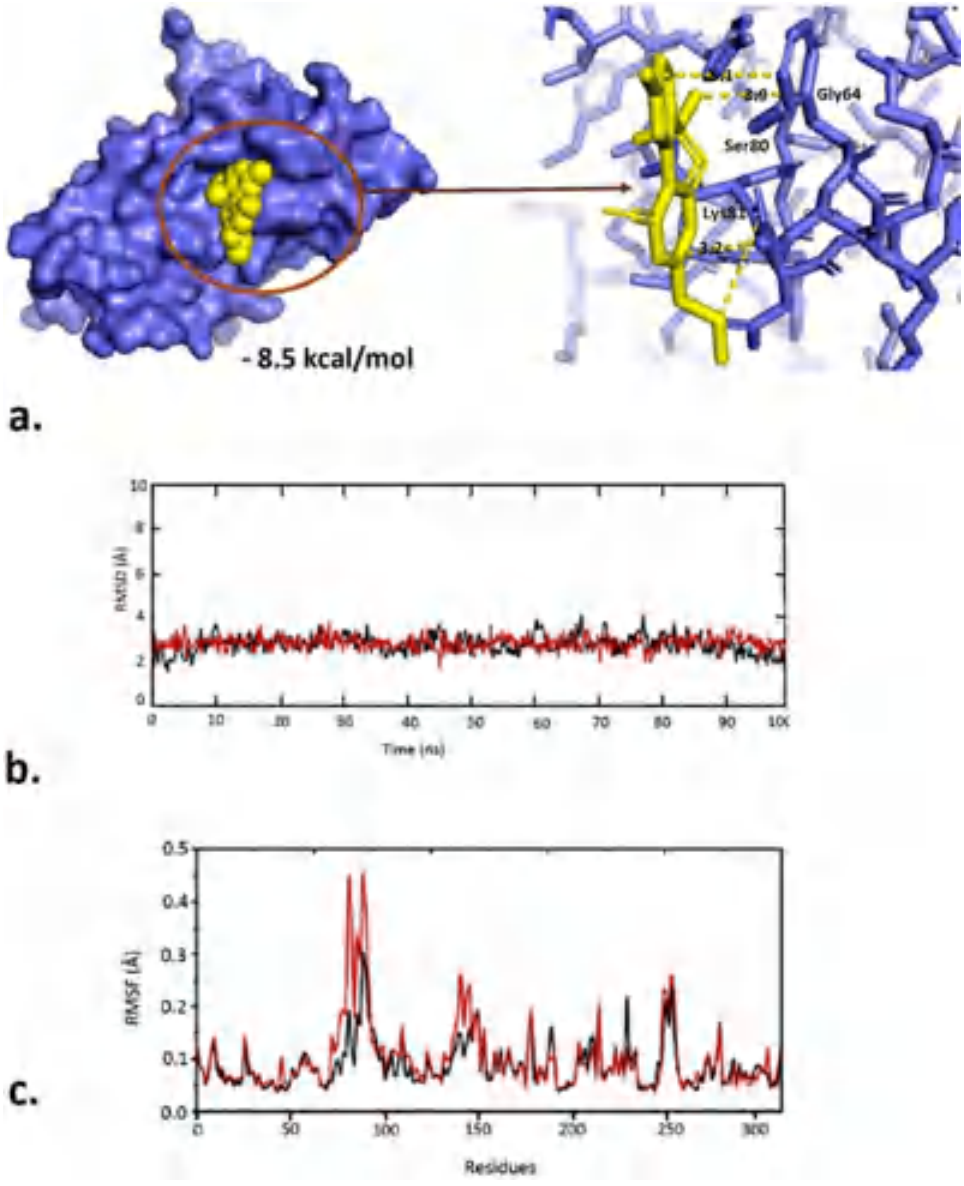

Figure 4. (a) Molecular docking of NRP1 protein with CVN. (b) RMSD plot comparison between the protein and docked complex. (c) RMSF plot comparison between the protein and docked complex. The red color indicates the protein and black indicates the docked complex.

Table 1. MM-GBSA calculation of all protein ligand complexes

Compound	ACE2	TMPRSS2	NRP1	interleukin (IL)-6
Cannabivarin	−50.23/−50.21	−48.24/−48.41	−31.02/−30.28	−31.03/−31.05
Cannabidiol	−43.2/−42.98	−31.47/−30.95	−34.28/−33.85	−37.12/−37.93

The first value before "/" is the value for the first simulation run and the value after "/" is from the second time simulation run.

the 3NPS-CVN complex showed a variation from 1.8-2.3. The backbone RMSD of 7B6P-CVN was found to be ranging from 2.5-3.7 over 100 ns time scale. This analysis (Figures 1b, 2b, 3b, and 4b) revealed that the protein-ligand complex did not distort the structural conformation of the actual protein to a larger extent strengthening the candidature of CBD and CVN as post-Covid-CNS treatment.

Root mean square fluctuation

RMSF plot is useful in depicting residues that have experienced major fluctuations during the molecular dynamics simulation (Muralidharan et al., 2020). RMSF for C-alpha atoms of each amino acid and was plotted against the number of residues. RMSF values for 6CS2, 1ALU, 3NPS and 7B6P over 100 ns time scale fluctuated from 1.3-3.2, 0.2-0.45, 0.1-0.4 and 0.1-0.42 respectively. The RMSF values of respective protein-ligand complexes i.e. 6CS2-CBD, 1ALU-CBD, 3NPS-CVN and 7B6P-CVN were 1.5-3.1, 0.2-0.4, 0.1-0.28, 0.1-0.3 respectively.

These plots revealed a similar fluctuation pattern for both proteins and protein-ligand complexes for all four simulation studies (Figures 1c, 2c, 3c, and 4c). We can predict the conformational stability of the protein-ligand complex if there are no considerable changes before and after MD simulations

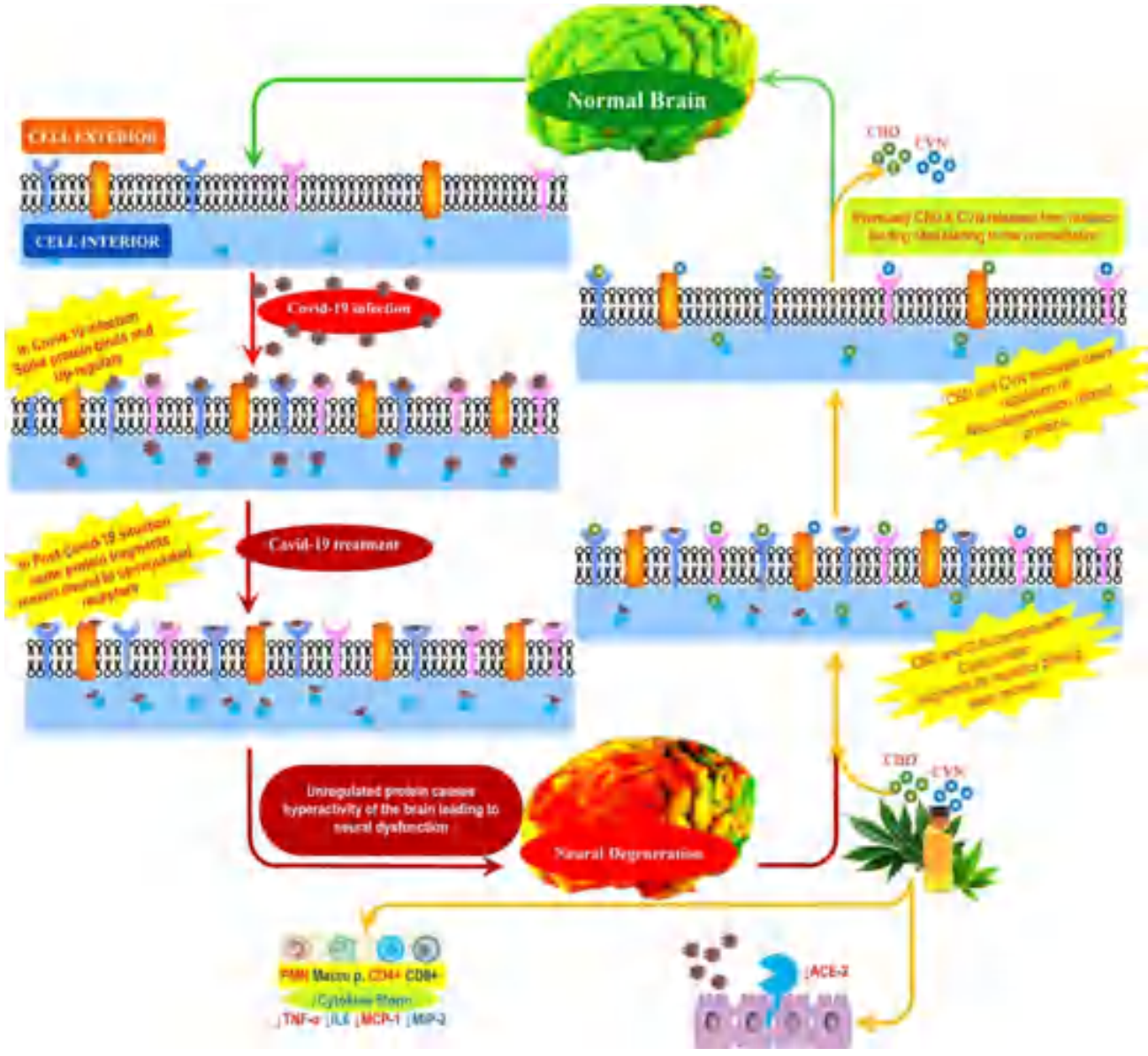

Figure 5. Schematic representation of the activity of CBD and CVN in post-Covid CNS treatment. Blue, Orange and Pink colored transmembrane protein represents ACE2 protein, TMPRSS2 and NRP1 protein respectively. The small blue-colored protein in the cell interior represents IL-6. Two-fold mechanism of CBD/CVN has been hypothesized over here. After Covid treatment the considered target proteins remain over expressed possible due to residual Covid proteins (designated by fraction of Covid virus). CBD/CVN cope with those binding sites of receptor proteins resulting into forced detachment of Covid residues from target proteins. After binding of CBD/CVN to receptors they downregulate them helping in reversal of neural homeostasis restoring normal brain functionality.

(Khan et al., 2020). In this study, 100 ns MD simulation of all four docked complexes revealed no major changes in the binding pattern. RMSF analysis indicated that binding of select ligands with considered proteins showed no major complications in terms of flexibility of protein and structural conformations thus reinforcing the effect of both CBD and CVN in post-Covid care.

Conclusion

The present outbreak of SARS-CoV-2, an influenza virus with neurotropic potential, emerged as neurological manifestations in a large proportion of the affected individuals as post-Covid symptoms throughout the world. Disorders of the central and peripheral nervous system have become common after the Covid treatment. People with these severe complications are most likely elderly with medical comorbidities, especially hypertension and other vascular risk factors. In this consequence, we tried to come up with some solutions to post Covid CNS symptoms. *In silico* screening of CBD and CVN from *Cannabis* has revealed potential therapeutic properties against these post-Covid neural complications. Moreover, some previous studies have focused on the effect of caffeine on *Cannabis* effectively. We have already started work to reveal the synergistic effects of CBD and caffeine in the post-Covid-CNS spasm as our future aspect of research. Till now, there is no specific treatment therapy for the post-Covid CNS syndrome hence we have focused on the post-Covid syndromes in this study. However, the present study is completely based on *in silico* screening and dynamic. The optimal dosage of these compounds has to be determined in the future through proper clinical trials.

Authors' contributions

AS, IS, GS conceived the idea. IS, GS and SB designed and executed the experiments and draft the manuscripts. IS executed major docking experiments, AS, MB and IS corrected the paper. All the authors have read and approved the article.

Acknowledgements

The financial help provided by the Department of Agriculture, Govt. of West Bengal, for establishing the Biswa Bangla Genome Centre is acknowledged. We also thank Dr. Stéphane Abel of Centre d'Etudes de Saclay for assisting in molecular dynamics and simulation of some Covid related proteins.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is partially funded by the Department of Biotechnology, Govt. of West Bengal.

ORCID

Indrani Sarkar http://orcid.org/0000-0002-5829-4510
Arnab Sen http://orcid.org/0000-0002-2079-5117

References

Al-Khafaji, K., Al-Duhaidahawi, D., & Taskin Tok, T. (2020). Using integrated computational approaches to identify safe and rapid treatment for SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 1–9.

Behl, T., Kaur, I., Bungau, S., Kumar, A., Uddin, M. S., Kumar, C., Pal, G., Shrivastava, K., Zengin, G., & Arora, S. (2020). The dual impact of ACE2 in COVID-19 and ironical actions in geriatrics and pediatrics with possible therapeutic solutions. *Life Sciences*, *257*, 118075. https://doi.org/10.1016/j.lfs.2020.118075

Beura, S., & Chetti, P. (2020). In-silico strategies for probing chloroquine-based inhibitors against SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, 1–13.

Bonaventura, A., Vecchié, A., Wang, T. S., Lee, E., Cremer, P. C., Carey, B., Rajendram, P., Hudock, K. M., Korbee, L., Van Tassell, B. W., Dagna, L., & Abbate, A. (2020). Targeting GM-CSF in COVID-19 pneumonia: Rationale and strategies. *Frontiers in Immunology*, *11*, 1625. https://doi.org/10.3389/fimmu.2020.01625

Cao, X. (2020). COVID-19: Immunopathology and its implications for therapy. *Nature Reviews. Immunology*, *20*(5), 269–270. https://doi.org/10.1038/s41577-020-0308-3

Clerkin, K. J., Fried, J. A., Raikhelkar, J., Sayer, G., Griffin, J. M., Masoumi, A., Jain, S. S., Burkhoff, D., Kumaraiah, D., Rabbani, LRoy., Schwartz, A., & Uriel, N. (2020). COVID-19 and cardiovascular disease. *Circulation*, *141*(20), 1648–1655. https://doi.org/10.1161/CIRCULATIONAHA.120.046941

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*, 42717. https://doi.org/10.1038/srep42717

Davido, B., Seang, S., Tubiana, R., & de Truchis, P. (2020). Post-COVID-19 chronic symptoms: A postinfectious entity? *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *26*(11), 1448–1449. https://doi.org/10.1016/j.cmi.2020.07.028

Durdağı, S. (2020). Virtual drug repurposing study against SARS-CoV-2 TMPRSS2 target. *Turkish Journal of Biology = Turk Biyoloji Dergisi*, *44*(3), 185–191. https://doi.org/10.3906/biy-2005-112

Isahq, M. S., Afridi, M. S., Ali, J., Hussain, M. M., Ahmad, S., & Kanwal, F. (2015). Proximate composition, phytochemical screening, GC-MS studies of biologically active cannabinoids and antimicrobial activities of Cannabis indica. *Asian Pacific Journal of Tropical Disease*, *5*(11), 897–902. https://doi.org/10.1016/S2222-1808(15)60953-7

Kar, P., Sharma, N. R., Singh, B., Sen, A., & Roy, A. (2020a). Natural compounds from Clerodendrum spp. as possible therapeutic candidates against SARS-CoV-2: An *in silico* investigation. *Journal of Biomolecular Structure and Dynamics*, 1–12.

Kar, P., Kumar, V., Vellingiri, B., Sen, A., Jaishee, N., Anandraj, A., Malhotra, H., Bhattacharyya, S., Mukhopadhyay, S., Kinoshita, M., & Govindasamy, V. (2020b). Anisotine and amarogentin as promising inhibitory candidates against SARS-CoV-2 proteins: A computational investigation. *Journal of Biomolecular Structure and Dynamics*, 1–11.

Khan, M. T., Ali, A., Wang, Q., Irfan, M., Khan, A., Zeb, M. T., Zhang, Y.-J., Chinnasamy, S., & Wei, D. Q. (2020). Marine natural compounds as potents inhibitors against the main protease of SARS-CoV-2. A molecular dynamic study. *Journal of Biomolecular Structure and Dynamics*, *2020*, 1–11. https://doi.org/10.1080/07391102.2020.1769733

Kovalchuk, A., Rodriguez-Juarez, R., Ilnytskyy, S., Li, D., Wang, B., Kovalchuk, I., & Kovalchuk, O. (2020). Fighting the storm: novel anti-TNFα and anti-IL-6 C. sativa lines to tame cytokine storm in COVID-19.

Lin, Q., Zhao, S., Gao, D., Lou, Y., Yang, S., Musa, S. S., Wang, M. H., Cai, Y., Wang, W., Yang, L., & He, D. (2020). A conceptual model for the outbreak of Coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, *93*, 211–216. https://doi.org/10.1016/j.ijid.2020.02.058

Marshall, M. (2020). How COVID-19 can damage the brain. *Nature*, *585*(7825), 342–343. https://doi.org/10.1038/d41586-020-02599-5

Mittal, L., Kumari, A., Srivastava, M., Singh, M., & Asthana, S. (2020). Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. *Journal of Biomolecular Structure and Dynamics*, 1–19.

Muralidharan, N., Sakthivel, R., Velmurugan, D., & Gromiha, M. M. (2020). Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 Protease against COVID-19. *Journal of Biomolecular Structure and Dynamics*, *16*, 1–6. https://doi.org/10.1080/07391102.2020.1752802

Palit, P., Chattopadhyay, D., Thomas, S., Kundu, A., Kim, H. S., & Rezaei, N. (2020). Phytopharmaceuticals mediated Furin and TMPRSS2 receptor blocking: Can it be a potential therapeutic option for Covid-19? *Phytomedicine*, 153396. https://doi.org/10.1016/j.phymed.2020.153396

Prajapat, M., Shekhar, N., Sarma, P., Avti, P., Singh, S., Kaur, H., Bhattacharyya, A., Kumar, S., Sharma, S., Prakash, A., & Medhi, B. (2020). Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein S1 domain and ACE2 interaction in SARS-CoV-2. *Journal of Molecular Graphics & Modelling*, *101*, 107716. https://doi.org/10.1016/j.jmgm.2020.107716

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics (Oxford, England)*, *29*(7), 845–854. https://doi.org/10.1093/bioinformatics/btt055

Raphael, K. G. (2020). Concerns raised by publication of Antonini et al., "Outcome of Parkinson Disease patients affected by Covid-19". *Movement Disorders: Official Journal of the Movement Disorder Society*, *35*(8), 1297. https://doi.org/10.1002/mds.28180

Russo, E. B., Guy, G. W., & Robson, P. J. (2007). Cannabis, pain, and sleep: Lessons from therapeutic clinical trials of Sativex, a cannabis-based medicine. *Chemistry & Biodiversity*, *4*(8), 1729–1743. https://doi.org/10.1002/cbdv.200790150

Sarkar, I., & Sen, A. (2020). *In silico* screening predicts common cold drug Dextromethorphan along with Prednisolone and Dexamethasone can be effective against novel Coronavirus disease (COVID-19). *Journal of Biomolecular Structure and Dynamics*, 1–5.

Sarma, P., Kaur, H., Kumar, H., Mahendru, D., Avti, P., Bhattacharyya, A., Prajapat, M., Shekhar, N., Kumar, S., Singh, R., & Singh, A. (2020). Virological and clinical cure in COVID-19 patients treated with hydroxychloroquine: a systematic review and meta-analysis. *Journal of Medical Virology*, *92*(7), 776–785.

Singh, D. D., Han, I., Choi, E. H., & Yadav, D. K. (2020). Immunopathology, host-virus genome interactions, and effective vaccine development in SARS-CoV-2. *Computational and Structural Biotechnology Journal*, *18*, 3774–3787. https://doi.org/10.1016/j.csbj.2020.11.011

Singh, D. D., Han, I., Choi, E. H., & Yadav, D. K. (2020a). Recent advances in pathophysiology, drug development and future perspectives of SARS-CoV-2. *Frontiers in Cell and Developmental Biology*, *8*, 580202. https://doi.org/10.3389/fcell.2020.580202

Tayyab, M., & Shahwar, D. (2015). GCMS analysis of *Cannabis sativa* L. from four different areas of Pakistan. *Egyptian Journal of Forensic Sciences*, *5*(3), 114–125. https://doi.org/10.1016/j.ejfs.2014.07.008

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Troyer, E. A., Kohn, J. N., & Hong, S. (2020). Are we facing a crashing wave of neuropsychiatric sequelae of COVID-19? Neuropsychiatric symptoms and potential immunologic mechanisms. *Brain, Behavior, and Immunity*, *87*, 34–39. https://doi.org/10.1016/j.bbi.2020.04.027

Wang, B., Kovalchuk, A., Li, D., Ilnytskyy, Y., Kovalchuk, I., & Kovalchuk, O. (2020). *In search of preventative strategies: Novel anti-inflammatory high-CBD cannabis sativa extracts modulate ACE2 expression in Covid-19 gateway tissues*.

Wijeratne, T., & Crewther, S. (2020). Post-COVID 19 Neurological Syndrome (PCNS); a novel syndrome with challenges for the global neurology community. *Journal of the Neurological Sciences*, *419*, 117179. https://doi.org/10.1016/j.jns.2020.117179

Yin, Y., Zhou, Z., Liu, W., Chang, Q., Sun, G., & Dai, Y. (2017). Vascular endothelial cells senescence is associated with NOD-like receptor family pyrin domain-containing 3 (NLRP3) inflammasome activation via reactive oxygen species (ROS)/thioredoxin-interacting protein (TXNIP) pathway. *The International Journal of Biochemistry & Cell Biology*, *84*, 22–34. https://doi.org/10.1016/j.biocel.2017.01.001

Zhang, M. Q., & Wilkinson, B. (2007). Drug discovery beyond the 'rule-of-five' . *Current Opinion in Biotechnology*, *18*(6), 478–488. https://doi.org/10.1016/j.copbio.2007.10.005

Archives of Microbiology (2022) 204:33
https://doi.org/10.1007/s00203-021-02635-6

ORIGINAL PAPER

Metagenomic outlooks of microbial dynamics influenced by organic manure in tea garden soils of North Bengal, India

I. Sarkar[1] · P. Kar[2] · G. Sen[3] · S. Chhetri[2] · M. Bhattacharya[3] · S. Bhattacharyya[4] · A. Sen[1,2,4]

Received: 23 June 2021 / Revised: 2 December 2021 / Accepted: 3 December 2021

Abstract

Soil microbial diversity consisted of both culturable and non-culturable microbes. The cultivated microbes can be identified by conventional microbiological processes. However, that is not possible for the non-culturable ones. In those cases, next-generation sequencing (NGS)-based metagenomics become useful. In this study, we targeted two very popular tea gardens of Darjeeling hills—Makaibari (Mak) and Castleton (Cas). The main difference between these two study areas is the type of manure they use. Mak is solely an organic tea garden using all organic manure and fertilizers whereas Cas uses inorganic pesticides and fertilizers. The main aim was to compare the effect of organic manure over chemical fertilizers on the soil microbiomes. We have performed the 16 s metagenomics analysis based on the V3–V4 region. Downstream bioinformatics analysis including reverse ecology was performed. We found that the overall microbial diversity is higher in Mak compared to Cas. Moreover, the use of organic manure has reduced the population of pathogenic bacteria in Mak soil when compared to Cas soil. From the observations made through the metagenomics analysis of Mak and Cas soil samples, we may conclude that the application of organic manure supports the population of good bacteria in the soil which may eventually impact the tea garden workers' health.

Keywords Tea garden · Soil metagenomics · Fertilizers · Stable ecotype model

Introduction

The microbial diversity of soil is enormous. Both culturable and non-culturable microbes enrich the soil microbial population. Some phylogenetic surveys on soil ecosystems made evident that the number of prokaryotic species present in a specific soil sample is far more than the known cultured prokaryotes (Daniel 2005). Conventional techniques for isolating and identifying the culturable microbes are not enough to study the overall diversity of soil microorganisms since it will exclude the considerable portion of non-culturables. Fortunately, with the advancement of metagenomic analysis, we can now gain a holistic idea about the microbial diversity of a specific soil sample.

Metagenomic analyses endow extensive information about the structure, composition, and predicted gene functions of varied environmental assemblages. Hence, 16 s metagenomic analysis based on the V3–V4 region has become a popular practice to unveil the effect of biotic and physicochemical factors on the overall microbial population of soil (Kakirde et al. 2010; Nesme et al. 2016).

There are some aspects of Metagenomics which are crucial to get the most accurate and relevant data from a sample. Extraction and purification of high-quality DNA is the major pre-requisite of any successful metagenomic study. The average insert size or the length of sequence reads for a high-throughput sequencing approach is also crucial. An appropriate metagenomics screening strategy should be adapted

I. Sarkar and P. Kar have contributed equally in this paper.

Communicated by Erko Stackebrandt.

✉ A. Sen
senarnab_nbu@hotmail.com

[1] Bioinformatics Facility, University of North Bengal, Siliguri, West Bengal, India

[2] Department of Botany, University of North Bengal, Siliguri, West Bengal, India

[3] Department of Tea Science, University of North Bengal, Siliguri, West Bengal, India

[4] Biswa Bangla Genome Centre, University of North Bengal, Siliguri, West Bengal, India

Published online: 19 December 2021

to address the specific question(s) of interest (Kakirde et al. 2010).

Despite all these obstacles, next-generation sequencing (NGS) based metagenomics provides us direct access to the uncultivated microbes. The high-throughput sequencing technology has equipped the field of microbiology with new phyla, class, genera, species as well as functional microbial genes (Nesme et al. 2016). The ability of metagenomics can answer numerous what, how who and why questions. For example, which type of soil hosts which kind of microbes? How the microbes interact with each other and also with the surrounding environment? Do they act synergistically or antagonistically? How the microbes act on changing environments and so on.

India is one of the major biodiversity hubs on this planet comprising of several different types of soils. This versatility ranges from snow soil to desert sand, from beach sands to most fertile riparian soil. The geographical location of this country has blessed it with a large area of agriculture friendly fertile soil with high yielding capacity. Paddy, wheat, green vegetables are common to grow in most of the Indian soils. However, one specific beverage crop that originally came to this country from outside and got very well adapted to the North-Eastern part of India is tea. Darjeeling tea has become world-famous for its brilliant aroma and color. Modern science has recognized the impact of the microbial population on the yielding capacity of the soil. Thus, it is now well known that soil along with its microbial communities can modulate not only the environment beneath the earth's crust but also above it including the crops that are grown in the soil along with other higher-order organisms and humans dependent on that soil in particular means.

In this study, we targeted two very popular tea gardens of Darjeeling hills—Makaibari (Mak) and Casselton (Cas). The main difference between them is the type of manure they use. Mak is solely an organic tea garden using all organic manure and fertilizers whereas Cas uses inorganic pesticides and fertilizers. Our main aim is to identify different sets of microbes that are present in these two tea garden soils using NGS-based metagenomic sequencing and to explore how the microbial population of both these tea gardens might be affected by the types of fertilizers being used. In this study, we deciphered the overall microbial population of both Makaibari (Mak) and Casselton (Cas) soil with special reference to their interaction among each other in terms of both complementation (synergy) or competition (antagonistic property). Along with that, we also tried to investigate whether the microbial population of two selected tea gardens may somehow affect the overall health quality of the tea garden workers or not.

Materials and methods

Field of study, sample collection, and soil testing

We have chosen two popular tea gardens from the Darjeeling hill region—Makaibari (26.8716° N, 88.2678° E) and Casselton (26.8659° N, 88.2777° E). The distance between these two tea gardens is only 12 min (4.0 km) via NH110 and they were on the same valley of the hill. Makaibari is solely an organic tea using organic fertilizers and manure whereas Casselton uses chemical pesticides and chemical fertilizers. Soil samples were randomly collected from the rhizosphere region of tea plants. Debris from the samples like roots, pebbles, etc. was removed by hand. Soil texture was assessed by the field method. The moisture percentage of soil samples was determined from the difference in weight of freshly collected and oven-dried soil samples. The clean air-dried samples were passed through a sieve and crushed with mortar and pestle. Soil pH, Electrical conductivity, and Loss of ignition were estimated following the protocol of (Baruah and Barthakur 1997). Other important parameters like organic carbon (Walkey and Black 1974), total soil Nitrogen (Jackson 1973) phosphorus as phosphate (Baruah and Barthakur 1997; Jackson 1973; Bray and Kurtz 1945) and potassium (Chapman and Pratt 1962), sulphur was determined during soil analysis. The level of micronutrients was qualitatively assessed by micronutrient kit. Information regarding the health of the tea garden workers were collected from a survey-based approach. The persons directly associated with the tea workers health of both organic and inorganic manure-based tea gardens were interviewed for collecting information regarding the present health scenario of those gardens.

DNA isolation

Soil DNA was isolated using rhizosphere soil of Makaibari and Casselton tea garden. Before initiating the isolation, 2 g of respective soil (three replicates per sample) was mixed with 4 ml of 1X Tris–EDTA buffer followed by proper vortexing (4–5 min) in 50 ml Oakridge tube. Before cell lysis, 150 µl of lysozyme (50 mg/ml), 100 µl of Proteinase K (20 mg/ml) and 600 µl of freshly prepared 10% SDS were mixed with the soil samples and the sample was incubated at 37 ºC for 90 min with gentle shaking every 15 min interval. After incubation, 1 ml 5 M sodium chloride, 1.6 ml CTAB/NaCl was mixed with respective solution and was further incubated at 65ºC for 30 min with occasional mixing in between to release the DNA from microbial cells. The supernatant containing microbial DNA was extracted with chloroform–isoamyl alcohol (24:1, v/v) and collected in a new tube after centrifugation at 6000 rpm for 15 min at

room temperature. The aqueous phase containing DNA was precipitated with 0.6 volumes of cold isopropanol and 0.1 volumes of 3 M sodium acetate followed by 2 h incubation at − 20 ºC. The DNA pellets were obtained by centrifugation at 10,000 rpm for 30 min at 4 ºC, washed with cold 70% ethanol, and dissolved in 100 µl of 1X Tris–EDTA buffer. To evaluate the purity of the extracted DNA, absorbance ratios at 260 nm/280 nm (DNA / protein) were determined and the DNA was sent for 16 s Metagenomics amplicon sequencing (V3–V4) to Genotypic Technology Pvt. Ltd.3.

PCR amplification of V3–V4 region of 16 s gene

About 40 ng of extracted DNA was used for amplification along with 10 pM of each primer (5' AGAGTTTGATG-MTGGCTCAG3' primer for forward sequence and 5' TTA CCGCGGCMGCSGGCAC3' primer for reversed sequence). The initial denaturation temperature was set as 95 °C. The denaturation was done for 15 s. Annealing was done at 60 °C for 15 s followed by elongation at 72 °C for 2 min. Final extension was done at 72 °C for 10 min. The final PCR product was stored at 4 °C. The amplified 16 s PCR Product was purified and subjected to GEL Check and Nanodrop QC. The Nano Drop readings of 260/280 at an approximate value of 1.8 to 2 were used to determine the DNA's quality.

Overall sequencing procedure

The amplicons from both samples were purified with Ampure beads to remove unused primers and an additional 8 cycles of PCR were performed using Illumina barcoded adapters to prepare the sequencing libraries. Libraries were purified using Ampure beads and quantitated using Qubit dsDNA High Sensitivity assay kit. Sequencing was performed using Illumina Miseq with 2 × 300PE v3 sequencing kit.

Processing of metagenomics data

Raw data QC was done using FASTQC and MULTIQC, followed by trimming of adapters and low-quality reads by TRIMGALORE. The trimmed reads were further taken for processing which includes merging of paired-end reads chimeria removal and OUT abundance calculation and estimation correction—this was achieved by Parallel-META pipeline. This workflow enabled highly accurate investigations at genus level. The databases used were SILVA (https://www.arb-silva.de/) / GREENGENES (https://greengenes.secondgenome.com/) and NCBI (https://www.ncbi.nlm.nih.gov/). Each read was classified based on % coverage and identity. A schematic diagram of the overall 16 s metagenomics process has been diagrammatically represented in Supplementary Fig. 1.

Reverse ecology analysis

Reverse ecology analysis is a simple yet effective way to study the interaction among microbes present in specific sample. This analysis considers both competition and complementation to assess the overall interaction among microbes.

The present metagenomic study identified microbes up-to genus level. We used a cut-off value of 200 sequence count i.e. if the count is less than 200 for a specific genus, we simply did not consider it. This was done purely to have a manageable amount of out-put data and to remove the possibilities of false-positives. We have used the same cut-off value for all further analysis in this study.

The whole genome sequences of the type strains from the identified genus (with count > 200) of both Mak and Cas were considered for reverse ecology analysis. Their KEGG Ontology (KO) information were retrieved from KAAS database. The KO information was fed into RevEcoR, an R-based package (Cao et al. 2016) to compute the competition and complementation indices among the studied strains.

Results

Physicochemical properties of Mak and Cas soil

Both the tea garden considered for this study had loam soil. Mak soil was light, friable loam with porous subsoil. This soil type is preferred for tea due to free percolation of water. The Cas soil was clay type. The low pH of both the soils indicated towards the acidic nature of the soil which is good for tea. Results of soil physicochemical analysis are shown in the table (Supplementary Table 1). The results were compared with soil physicochemical standards recommended by the Tea Board of India. The clayey soils of the tea plantations have low pH, sulphur but high organic carbon, organic matter, total nitrogen and P2O5. K2O was optimum in soil collected from Mak but high in soil from Cas. Micronutrients like boron, manganese, zinc, and copper were low while Molybdenum was moderate in both the plantations. Iron was optimum in Mak but low in Cas. The soil types of the tea gardens were not largely different. However, the difference arose in the fertilizers used by these two tea gardens. Mak is practicing with the organic manure filled with vermicompost, bio-fertilizers, and organic manure while Cas is totally dependent upon the inorganic manure and pesticides for maintaining sodium (Na):pottasium (K) ratio, weed control along with pest and disease management.

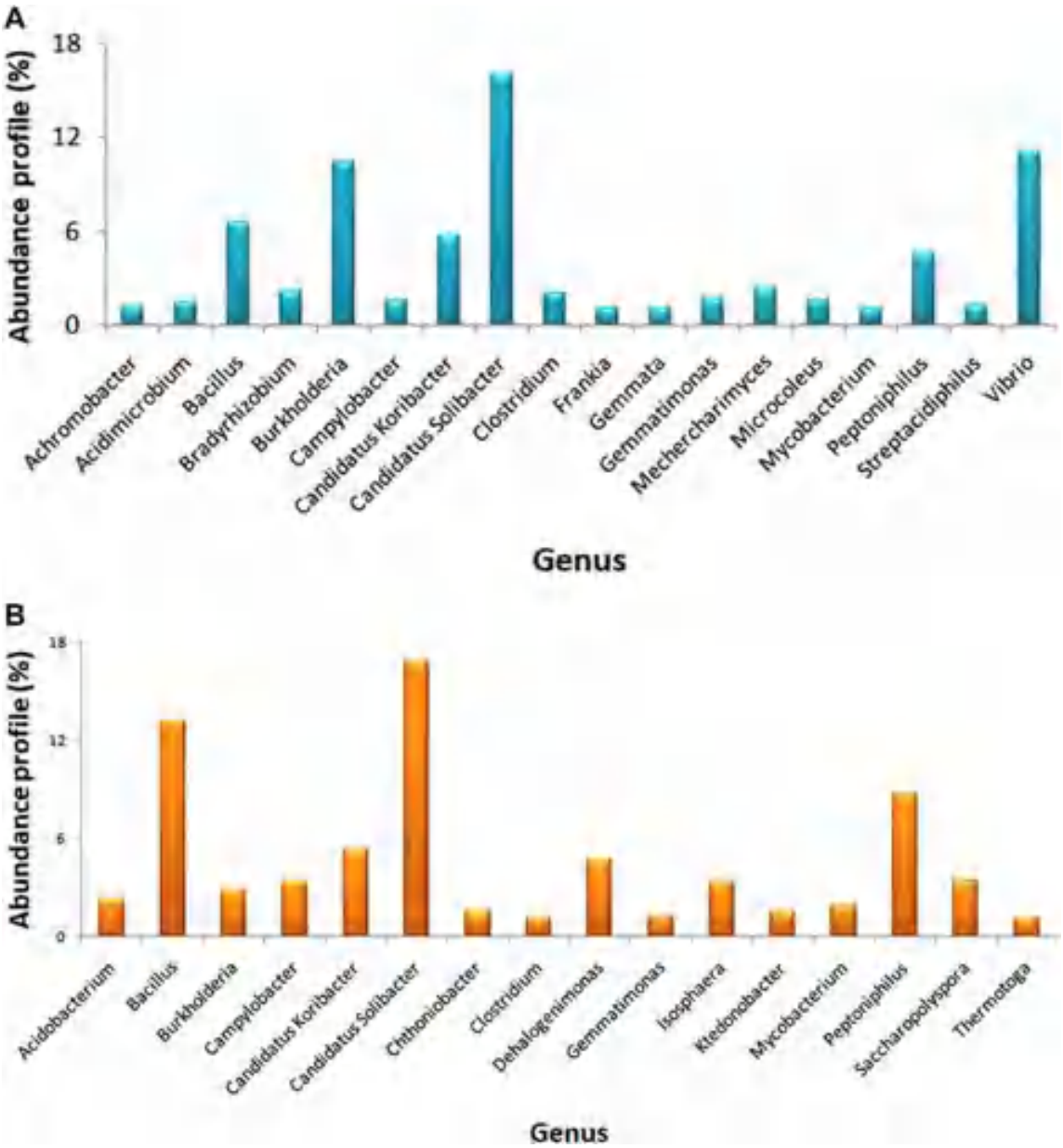

Fig. 1 (**a**) Microbial abundance profile of Mak. (**b**) Microbial abundance profile of Cas (color figure online)

Primary data summary

The paired-end reads from Mak and Cas soils gave 56% and 55% average GC, respectively. There were 0.2 M sequences for each read of Cas with 67.95% duplication value and 0.3 M sequences for each read of Mak with 72% duplication value. The read lengths were 256 bp and 205 bp for forward and reverse sequences of Cas samples. Mak forward and reverse sequence lengths were 258 bp and 189 bp, respectively. FastQC report revealed good quality reads indicating successful metagenomic sequencing. The 16 s metagenomics data for Mak and Cas has been submitted in NCBI SRA under the BioSample accession: SAMN21875714 with BioProject ID: PRJNA766783.

MAK is more populated than CAS with good bacteria leading to a stable ecotype model

The taxonomic abundance profiling identified the microbial abundance from phylum to genus level. It was found that both soils shared a large set of bacteria however, their relative abundance was not the same. For instance, *Cyanobacteria* and *Gemmatimonadetes* were present in Mak constituting 1.32% and 1.24% of the total microbial population,

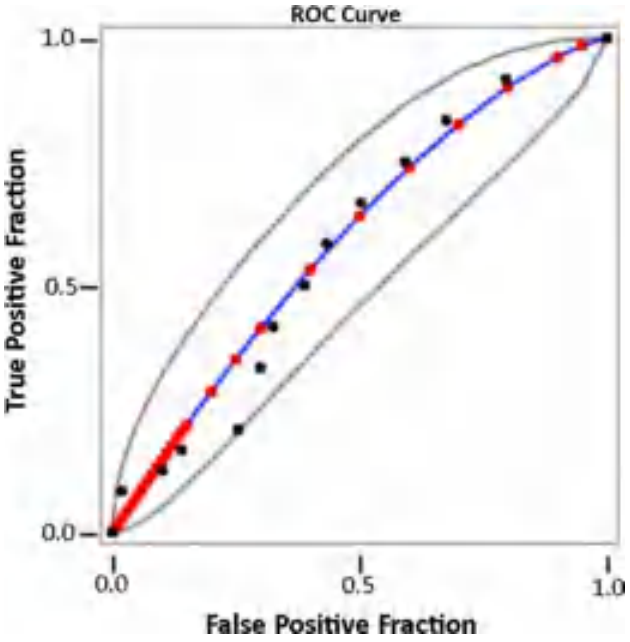

Fig. 2 The receiver operating characteristic (ROC) plot analysis among Mak and Cas (color figure online)

respectively, whereas they constituted less than 0.5% in Cas soil. The major microbial phyla identified in both soil samples were *Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Choloflexi*, *Firmicutes*, *Bacteroidetes*, *Verrucomicrobia*, *Planctomycetes* (Fig. 1a, 1b). We did the one-way ANOVA test, however, found no significant difference among the overall microbial population of Mak and Cas. The f ratio value was found to be 0.00589 and p value was 0.939465 and it was not significant even at $p < 0.01$. The receiver operating characteristic (ROC) curve (http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html) was plotted based on the overall microbial diversity among Mak and Cas (Fig. 2). The analysis gave fitted AUC (area under ROC) as 0.59 and empiric AUC was 0.585 suggesting less discrimination between the overall microbial population of Mak and Cas.

However, a distinct pattern was observed between the two soil samples when only pathogenic microbes were considered. For instance, the relative abundance of the pathogenic microbial population was found to be more in Cas than Mak (Fig. 3). The abundance profile of pathogenic microbes like *Burkholderia, Campylobacter*, and *Bacillus* were much higher in Cas than in Mak. Moreover, the abundance of *Mycobacterium* was also more in Cas than Mak. Along with those mentioned genera, the presence of *Candidatus Solibacter, Candidatus Koribacter, Peptoniphilus, Peptoniphilus, Clostridium, Gemmatimonas* were found in both Cas and Mak. Their abundance was not very high in any of the soil samples. *Bradyrhizobium, Microcoleus, Acidimicrobium, Streptacidiphilus, Achromobacter, Gemmata and Frankia* were solely present in Mak but not in Cas however, *Dehalogenimonas, Saccharopolyspora, Isosphaera, Acidobacterium, Chthoniobacter, Ktedonobacter, Thermotoga* were solely present in Cas but not in Mak. This indicated the differential bacterial population among Cas and Mak. When the one-way ANOVA test was performed the f ratio came to be 7.75 with p value 0.010285 and the result was significant at $p < 0.05$ (at $p < 0.01$ the difference was non-significant). A PCA plot based on the pathogenic microbial population also supported the ANOVA results where Mak and Cas were placed in two different quadrants of the PCA plot (Fig. 4).

To find the species diversity between the two soil samples α diversity of both samples were exploited. Alpha (α) diversity is a direct measure of mean species diversity of habitat and a higher α diversity value indicates more diversity. The α-diversity value of Cas was 48.69 and for Mak it was 56.62 pointing to more species richness in Mak. Rare-fraction curve that allows us to calculate the species richness from a given number of individual samples was further implemented to support our aforementioned hypothesis. A common pattern of this curve is, it grows rapidly at first due to the most common species present in the samples and gradually becomes a plateau as the rarest species remain to be sampled. In Cas, the curve started to get a plateau state at species count 1400 (Fig. 5a) wherein Mak the stage came at species count 3000 (Fig. 5b). Hence, it is evident from taxonomical abundance profiling, α diversity and rare-fraction curve analysis that, Mak is more ecologically diverse with a higher microbial population rather than Cas.

Reverse ecology analysis was implemented to get a birds-eye view on the complex microbial interaction and signaling network going on within Mak and Cas soil samples. The reverse ecology analysis revealed that, the complementation index for Mak microbial population varied from 0.79–0.97 and that of Cas was 0.68–0.85. The competition index for both the samples were considerably low (0.21–0.39 for Mak and 0.32–0.41 for Cas). The differences between complementation and competition in both Mak and Cas were statistically significant (t test at $p < 0.001$). Moreover, the complementation among Mak population was more than Cas population ($p < 0.001$). This also supports that, the Mak microbial population has formed a more stable ecotype model than Cas which indicates a better natural selection forming and maintaining specific genetic clusters.

Discussion

Microbial population of Mak formed a stable ecotype model

We have found a clear picture about the microbial diversity of both studied tea gardens. Here are the major findings we

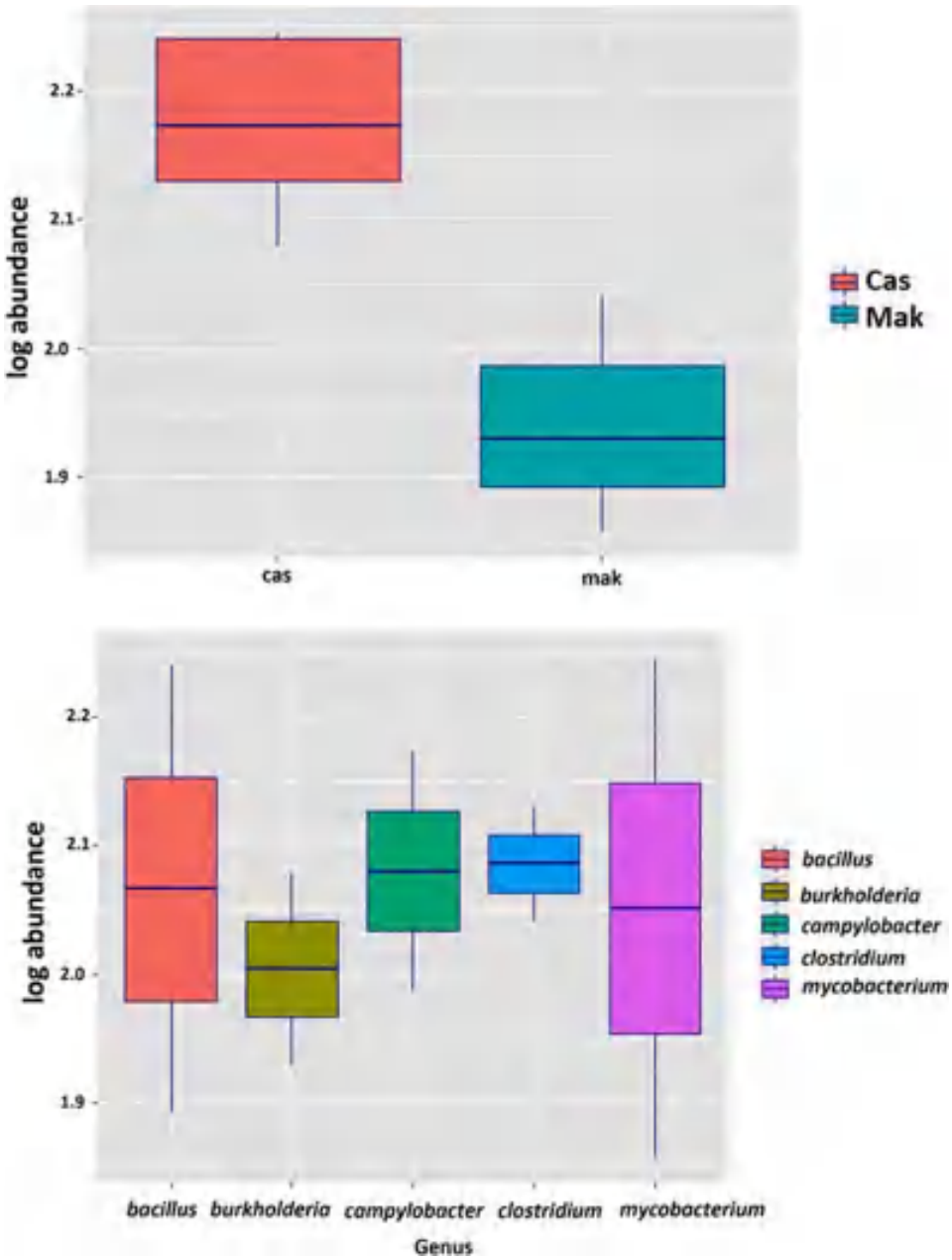

Fig. 3 Comparative pathogenic microbial abundance profile of Mak vs Cas (color figure online)

got (a) the overall soil physicochemical properties were alike as they belong to the same eco-geographical region and altitudinal level. (b) The overall bacterial diversity was more in Mak than Cas. (c) Moreover, Cas population contained more pathogenic genus than Mak. This clearly indicated a positive effect of organic manure in comparison to inorganic/chemical fertilisers. (d) The complementation values (obtained from reverse ecology analysis) among the Mak population was higher than Cas population. This may indicate a stable ecotype model (SEM) (Shapiro and Polz 2015) persisting in Mak where the main carbon source of soil is organic manure. It is a well-known fact that fertilizers have a direct impact on soil microbial population playing a pivotal role in both biogeochemical cycling and ecological processes (Li et al. 2017). Certain microbial taxa display ecological coherence in response to environmental variables. Based on substrate preference and life strategies, those microbes can be grouped into r-selected or k-selected categories. However, it is difficult to gain such knowledge at a lower taxonomic level (genus or species level).

Use of organic manure increased the microbial diversity of Mak

It has been documented previously that, continuous exposure of fertilization (both organic and inorganic) leads to the

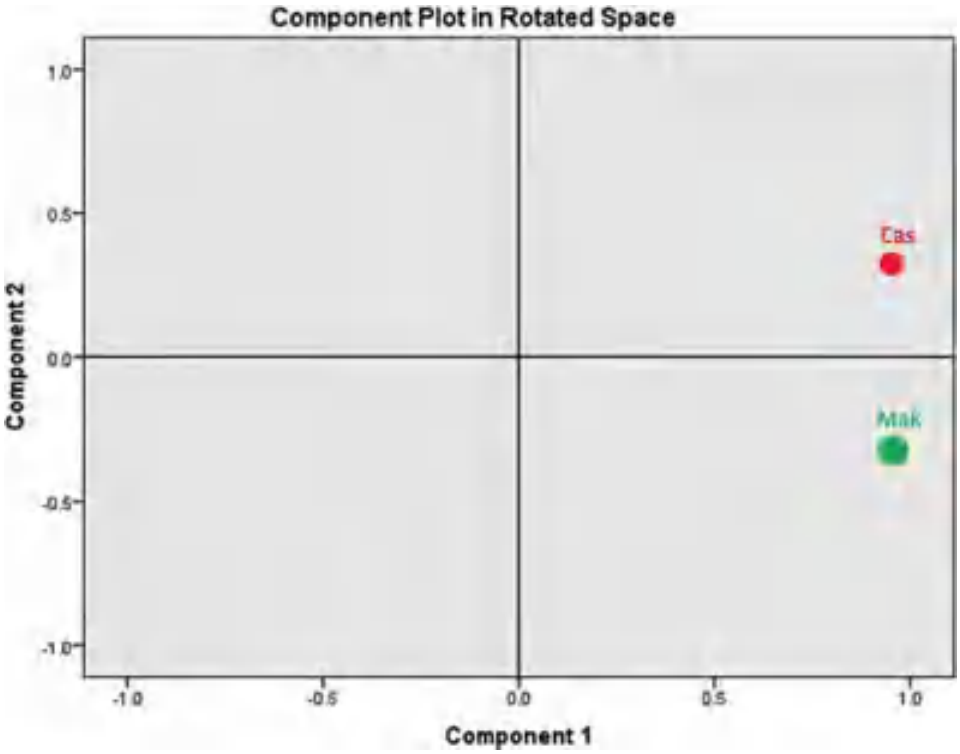

Fig. 4 PCA plot analysis of pathogenic microbial abundance profile of Mak vs Cas (color figure online)

addition of a specific category of carbon (C) and nitrogen (N) source to the soil. Over a period of time, a set of bacteria, capable to handle those specific C and N sources will proliferate in that agricultural field. This practice in the long run is good for providing agroecosystem stability. Organic manures are composed of different decomposing materials hence contain diverse C and N sources. On the contrary, chemical fertilizers are always well defined with their source of C and N (Li et al. 2017). As a result, it may well be predicted that a field exposed to long-term organic manure will house a more versatile microbial population utilizing various kinds of nutrient sources than a field exposed to defined inorganic manure. Makaibari (Mak) tea garden is popular for using organic manure since its inception whereas, Castleton (Cas) uses inorganic fertilizers. These differential practices are thus, playing a major role in the microbial population between these two tea gardens.

Microbial populations of tea garden soil bear a relation with tea garden worker health

In the tea plantation sector, safety and security issues of workers are overlooked widely (Roy Chowdhury et al. 2014). Most of the workers are ignorant about the consequences of the exposure to chemicals, environmental factors, etc. Extensive use of chemical fertilizers and pesticides results in the degradation of soil and water bodies. Agricultural chemical inputs gain access into human body systems through three major means: (1) oral ingestion, (ii) infiltration through the skin, and (iii) breathing (Roy Chowdhury et al. 2014; Rodríguez-Eugenio et al. 2018; Rajput et al. 2021; Bottone 2010; Ahmmed and Hossain 2016; http://www.tezu.ernet.in › project reports). Previous studies on tea garden workers showed the prevalence of neurological, gastrointestinal, renal and hepatic toxicity (Inglis and Sagripanti 2006; Frost 2001; Picard et al. 2005; Bae et al. 2002; Chenoll et al. 2015) among them. Most of the tea garden workers are prone to respiratory ailments such as tuberculosis and skin disorders (Gayathri and Arjunan 2019). Since the vast majority of workers in the tea plantation are women, concerns have centered on the potential reproductive hazards of chemical exposure and their impact on pregnant women, nursing mothers within their lactation period, and their children (Rajbangshi and Nambiar 2020). The difference in the soil micro-flora of Mak in comparison with Cas from our study indicated the role of fertilizers and chemicals in the development of soil microrganisms. The abundance of beneficial flora in Mak may provide a positive effect on the health aspect of tea garden workers further directed towards the advantages of organic manure over chemical fertilizers.

Conclusion

Human–microbe affiliation establishes even when the person is there in the mother's womb and the connection between human and soil microbes launches before it starts walking on the ground. It is believed that soil microbes contribute

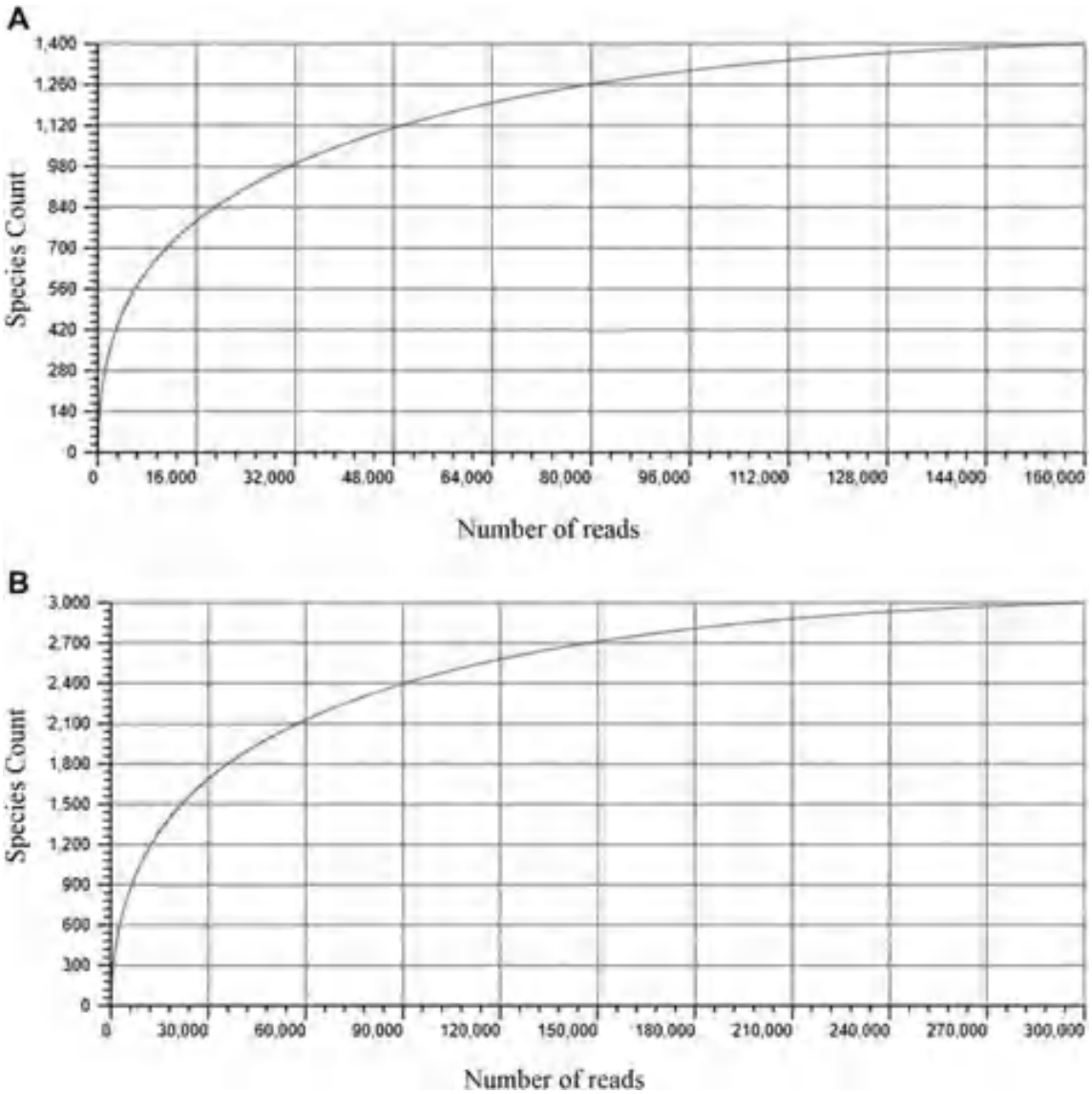

Fig. 5 Rare-fraction curve of (**a**) Mak and (**b**) Cas soil samples (color figure online)

considerably in developing the gut micro-flora and shape the overall human health. Soil dwellers, mainly soil micro-flora, play a paramount role in maintaining the biodiversity of a micro-habitat like tea gardens, paddy fields etc. The cultivable soil microbes are relatively easy to study while a large portion of culture-independent microbiomes remain largely illusory. In this consequence, soil metagenomics has become an important tool in studying the non-cultivable microorganisms present in a specific niche. In this present study, we did 16 s metagenomics of Makaibari (Mak) and Castleton (Cas) tea gardens from the Darjeeling region of India. The main difference between these two gardens is, Mak is an organic manure-based tea garden whereas Cas uses chemical fertilizers. Metagenomics revealed higher bacterial diversity in Mak than Cas. The pathogenic bacterial population was more in Cas than Mak indicating the positive feedback effect of organic manure on the overall bacterial population of soil. We investigated interactions among the identified genus from both Mak and Cas. A stable ecotype model was evident in Mak where microbes were showing synergistic effect (complementation) whereas in Cas soil, competition was more among the bacterial population revealing volatility of the ecosystem. Finally, the number of human pathogens was more in Cas than Mak which supported the better tea garden worker health report in Mak over Cas. Literature survey, as well as our own survey also supports this fact. Thus, this study indicates that organic fertilizers have a positive effect on the soil microbial population in particular and human health in general in that region.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00203-021-02635-6.

Acknowledgements We acknowledge Makaibari and Castleton tea gardens for providing us samples and other supports. Especial acknowledgement to Dr. P.K. Dutta, North Bengal Medica Nursing Home, Siliguri for sharing his knowledge about the health aspect of tea garden workers.

Author contributions AS conceived the idea and did the experimental design. PK, GS, SC, MB collected and prepared samples for metagenomics. MB, PK and SB did the soil analysis related work. AS, GS and IS did the bioinformatics analysis. Figures and art works are mostly done by IS. All the authors contributed in manuscript writing and approved.

Declarations

Conflict of interest The authors declare that the research paper was written in the absence of any commercial or financial relationships that could be construed as real or potential conflict of interest.

Research involving human and animal participants No animal or human were treated as sample in this study.

References

Ahmmed F, Hossain MI (2016) A study report on working conditions of tea plantation workers in Bangladesh. International Labour Organization.

Bae E-A, Han MJ, Song M-J, Kim D-H (2002) Purification of Rotavirus Infection-Inhibitory Protein from *Bifidobacterium* Breve K-110. COREE, REPUBLIQUE DE, Korean Society for Applied Microbiology, Seoul

Baruah TC, Barthakur HP (1997) A text book of soil analysis. Vikas Publishing House Pvt Ltd., New Delhi

Bottone EJ (2010) Bacillus cereus, a volatile human pathogen. Clin Microbiol Rev 23(2):382–398. https://doi.org/10.1128/CMR.00073-09

Bray RH, Kurtz LT (1945) Determination of total, organic, and available forms of phosphorus in soils. Soil Sci 59(1):39–46

Cao Y, Wang Y, Zheng X, Li F, Bo X (2016) RevEcoR: an R package for the reverse ecology analysis of microbiomes. BMC Bioinform 17(1):1–6. https://doi.org/10.1186/s12859-016-1088-4

Chapman HD, Pratt PF (1962) Methods of analysis for soils, plants and waters. Soil Sci 93(1):68

Chenoll E, Rivero M, Codoñer FM, Martinez-Blanch JF, Ramón D, Genovés S et al (2015) Complete genome sequence of *Bifidobacterium longum* subsp. infantis Strain CECT 7210, a probiotic strain active against rotavirus infections. Genome Announc 3:e00105-e115. https://doi.org/10.1128/genomea.00105-15

Daniel R (2005) The metagenomics of soil. Nat Rev Microbiol 3(6):470–478. https://doi.org/10.1038/nrmicro1160

Frost JA (2001) Current epidemiological issues in human campylobacteriosis. J Appl Microbiol 90:85S-95S. https://doi.org/10.1046/j.1365-2672.2001.01357.x

Gayathri MP, Arjunan R (2019) Health afflictions of tea plantation workers in Coonoor. Nilgiris Int J Health Sci Res 9(11):85–90

Inglis TJ, Sagripanti JL (2006) Environmental factors that affect the survival and persistence of *Burkholderia pseudomallei*. Appl Environ Microbiol 72(11):6865–6875. https://doi.org/10.1128/AEM.01036-06

Jackson ML (1973) Soil chemical analysis. Prentice Hall of India Private Limited, New Delhi, p 498

Kakirde KS, Parsley LC, Liles MR (2010) Size does matter: application-driven approaches for soil metagenomics. Soil Biol Biochem 42(11):1911–1923. https://doi.org/10.1016/j.soilbio.2010.07.021

Li F, Chen L, Zhang J, Yin J, Huang S (2017) Bacterial community structure after long-term organic and inorganic fertilization reveals important associations between soil nutrients and specific taxa involved in nutrient transformations. Front Microbiol 8:187. https://doi.org/10.3389/fmicb.2017.00187

Nesme J, Achouak W, Agathos SN, Bailey M, Baldrian P, Brunel D, Frostegård Å, Heulin T, Jansson JK, Jurkevitch E, Kruus KL (2016) Back to the future of soil metagenomics. Front Microbiol 7:73. https://doi.org/10.3389/fmicb.2016.00073

Picard C, Fioramonti J, Francois A, Robinson T, Neant F, Matuchansky C (2005) Review article: bifidobacteria as probiotic agents—physiological effects and clinical benefits. Aliment Pharmacol Ther 22:495–512. https://doi.org/10.1111/j.1365-2036.2005.02615.x

Rajbangshi PR, Nambiar D (2020) "Who will stand up for us?" the social determinants of health of women tea plantation workers in India. Int J Equity Health 19(1):1–10. https://doi.org/10.1186/s12939-020-1147-3

Rajput P, Sinha RK, Devi P (2021) Current scenario of pesticide contamination in water. In: Ahamad A, Siddiqui SI, Singh P (eds) Contamination of water. Academic Press, pp 109–119. https://doi.org/10.1016/B978-0-12-824058-8.00032-3

Rodríguez-Eugenio N, McLaughlin M, Pennock D (2018) Soil pollution: a hidden reality. FAO, Rome

Roy Chowdhury D, Paul M, Banerjee SK (2014) Review on the effects of biofertilizers and biopesticides on rice and tea cultivation and productivity. Int J Sci Eng Technol 2:96–108

Shapiro BJ, Polz MF (2015) Microbial speciation. Cold Spring Harb Perspect Biol 7(10):a018143. https://doi.org/10.1101/cshperspect.a018143

Walkey A, Black CA (1974) Critical examination of rapid method of determining organic carbon in soil. Soil Sc 1974(63):251

Springer Protocols:**Methods in Acttinobacteriology**; Humana Press
Editor: Dhanasekaran Dharumadurai

Methods for whole-genome analysis of Actinobacteria through Bioinformatics approaches

I Sarkar, G Sen and A Sen*

Bioinformatics Facility, Department of Botany, University of North Bengal

***Corresponding Author- senarnab_nbu@hotmail.com**

Abstract

Actinobacteria is one of the most diverse groups of bacteria inhabiting virtually all different niches. Actinobacteria can be found in plants, animals, soil, water and extremophile situations like desert soil, arctic soil and thermophilic condition. Next-generation Illumina sequencing provided us the whole genome sequences of microorganisms including Actinobacteria. Genomics study on Actinobacteria has availed us with vast knowledge regarding the interaction of this group with its respective niche. In this chapter, we will discuss several important and interesting Bioinformatics techniques vastly used for microbial research.

Key Words: Actinobacteria, Codon Usage, Bioinformatics, Sequencing, Phylogeny

Introduction

Background of Actinobacteria

Actinobacteria which was previously known as ‘actinomycetes’ or ray fungi form an important constituent of the microbial biome. It comes next to proteobacteria in terms of number and distribution. Actinobacteria are mostly aerobic, gram-positive to gram variable with high G+C content and occupya diverse microbial niche. They share some characteristics with fungi, such as colony morphology, mycelial growth and musty smell, on the other hand, peptidoglycan cell wall structure is common with bacteria. Phylogenetic studies based on 16S rRNA classifies actinobacteria into six classes i.e. Acidimicrobia, Corniobacteria, Nitriliruptoria, Rubrobacteria, Thermoleophilia and Actinobacteria (1).

Actinobacteria are widely distributed in various biotopes such as soil, water, permafrost, mammals, arthropods, plants etc. (1). Furthermore, various lifestyles are encountered among *Actinobacteria*, and the phylum includes pathogens (e. g. *Mycobacterium* sp. *Nocardia*sp. *Tropherym*asp. *Corynebacterium*sp. and *Propionibacterium* sp.), soil inhabitants (*Streptomyces* sp.), plant commensals (*Leifsonia*sp.), nitrogen-fixing symbionts (*Frankia*), and gastrointestinal tract (GIT) inhabitants (*Bifidobacterium*) (2).

Actinomycetes identification in conventional systems is difficult, but it can be done by performing biochemical tests (3). It can also be done by advanced methods such as utilizing software 'Actinobase' for genus-level identification, using image files. Phylogenetic relationship determination is made easy by 16S rRNA studies, up to species level recognition using Blast search.

Biological research has undergone tremendous progress after the advancement of genome sequencing. Genomic database mining provides us with an opportunity in studying the genome profiles of different organisms and also helps in studying comparative genomics. The ongoing proliferation of whole-genome sequences is a stepping stone for systems biology (4), which aims to study the integrated network constituted by the complete repertoire of genes (genome), the population of transcripts (transcriptome),the population of proteins (proteome), the population of metabolites (metabolome),and fluxes of an organism or cell, concerning intrinsic and environmental stimuli (5). Bioinformatic tools become relevant in this aspect. The language of DNA which is the four-letter alphabet that is expressed as triplet codon for an amino acid is the key for fundamental gene expression. The degeneracy of codons for 18 amino acids except methionine and tryptophan showed that different sequences of DNA produce identical protein sequences. The degeneracy mainly occurred in the third position of the codon. Data from whole-genome sequences help to study the preference of codons among organisms.It has been noticed that variation of choice of codons to represent amino acids is not only observed among species from the different taxonomic groups but also showed significant variation among individuals of the same species, across different genes in the same genome and even across regions in the same gene. But, the codon bias is most prominent in species from different taxonomic groups even in proteins with identical functions. This phenomenon of species-specific codon choice is known as "codon dialect," which signifies the codon-usage bias observed across different organisms (6).

Bioinformatics in Genomic Research

Through the development of algorithms and statistical testing, research can be carried out faster and more accurately. Bioinformatics is used in different fields of research; however, it is especially important in genomics, such as in genome analysis, gene identification, genome-wide association studies and evolutionary studies. Traditional and next-generation processes aim to sequence the genome allowing the

analysis of DNA sequences. However, these methods produce many fragments of DNA, like fragments of a jigsaw puzzle, which need to be aligned and compiled to create a final complete sequence. The use of bioinformatics can align these fragments quickly and cheaply, aiding genomic sequencing.The human genome was initially sequenced between 1990 and 2003 and has since been uploaded online and extensively annotated. Annotation is the process whereby genes and their protein products are labeled directly onto the genome. The volume and complexity of the produced data would have taken many years to compile manually. However, with the advent of bioinformatics, scientists can carry out the compilation and annotation processes quickly and with better precision.

Bioinformatics and identification of mutations

Bioinformatics is vital in the research of*de novo*mutations. One example of a method that is used to identify these mutations is whole-exome sequencing. Whole exome sequencing is used to sequence only the protein-coding regions of DNA (the exomes), which makes up only 1% of the genome, thereby making it much faster than genome sequencing. However, large quantities of data are produced whereby bioinformatics application becomes vital for data curation, sequence alignment, and analysis. Using whole-exome sequencing and bioinformatics, 50% of rare disease genes have so far been identified, with the rest is expected to be sequenced by 2020. Another use of bioinformatics is in the identification of cancerous mutations. Through the development of automated systems, large volumes of sequential data can be produced and used to identify previously unknown point mutations. Bioinformatics also works to create new algorithms that can compare different sequences, thereby aiding in the identification of mutations.

Bioinformatics and genome-wide association studies

Genome-wide association studies (GWAS) carry out genomic scans inan attempt to identify specific markers that can indicate an individual's susceptibility to a genetic disease. Genetic association between a specific marker and the disease can improve detection and treatment. If used on a large scale, this can also aid in the development of prophylactic treatments. To carry out GWAS, the genomes of individuals with a disease and those without a disease are compared. The development of highly automated systems has led to the high-throughput identification of single nucleotide polymorphisms (SNPs).By comparing SNPs, those which are more common in individuals with the disease can be identified and used as disease markers. This information is then stored online and made available to scientists across the globe. The first published GWAS was age-related macular degeneration (AMD). Out of 116,204 SNPs that were genotyped, one study observed a link between the complement factor "H" (CFH) gene and AMD. Therefore, individuals susceptible to AMD can be screened for the presence of the CFH gene.Several

other disease genes have been characterized after that to help doctors and other health care professionals in identifying the possible risk of a genetic disease and allowing for appropriate disease management.

Bioinformatics and evolutionary studies

By studying the changes in DNA within organisms and comparing them to other species, the genetic changes associated with evolution can be classified. Evolution is the process that involves small, cumulative changes in DNA that eventually leads to the formation of novel species.Bioinformatics has aided research in the evolutionary process by allowing comparison of DNA sequences, sharing of data, prediction of future evolution and classification of complex evolutionary processes.When put together, the data can be used to create a phylogenetic tree that can trace several species to their original ancestry. These are only a few of the myriad applications of bioinformatics within genetics. Overall, bioinformatics has thrown open enormous opportunities in the field of genomics and targeted gene therapy.

Materials

1. Fasta Nucleic acid and Fasta amino acid sequences can be downloaded from IMG database or NCBI.

2. Sequences for 16s, 23s, Multi –Locus sequences (AtpA, DnaB, GyrA, FtsZ, SecA) sequences in fasta format is required for phulogeny analysis

3. Software mentioned here: CodonW, DAMBE, ClustalW, MEGA, Perl.

4. Statistical anysis can be done through SPSS.

Methods

Codon Usage

The codon usage pattern is a unique feature of a particular organism. It helps us to understand gene expression, horizontal gene transfer and also enablesus to determine phylogenetic relationships between organisms. The study of codon usage patterns of several genes and genomes is a popular technique to characterize and analyze genomic trends from a bioinformatics-based perspective. Codon usage patterns and preferences vary significantly within and between organisms (1,4).

CODON W software developed by Pedan,1999 became very popular and widely used for studying codon usage and multivariate analysis because of its error-free analysis. The parameters such as GC content

(amount of guanine-cytosine in the nucleotide sequences),GC3 content(frequency of either G or C nucleotides in the third position of synonymous codon), the effective number of codons used in a gene(Nc), frequency of optimal codons(Fop),CBI (codon bias index),GRAVY(hydrophobicity of amino acids) are included in this analysis.The most obvious factor that determines codon usage is a mutational bias that shapes genome GC composition. Mutational bias is responsible not only for the intergenetic difference in codon usage but also for codon usage bias within the same genome.Most of the organisms with a balanced AT/GC genome have codon heterogeneity(Sen et al.,2007).Highly expressed genes contain a higher percentage of translationally optimal codons (7). Codon heterogeneity in the genome can be studied by GC content, GC 3 content, effective number of codons (Nc). Nc measures the overall codon bias of synonymous codons (8). It ranges from 20(in the case of one codon for one amino acid) to 61(where all codons are used).

Whole-genome sequences of actinobacteria can be obtained from the IMG database (www.img.jgi.doe.gov) using their respective genome accession numbers. Nucleotide sequences are downloaded for the codon analysis.For codon analysis, the CODONW version1.4.2 and the selected genome is put together in a folder. Before running the program, file extension 'fna' has to be changed into 'dat' format. Several parameters such as GC content, GC3 content, Nc, CBI, Fop, GRAVY, RSCU, and aromaticity are obtained as output. Data can be saved in excel format for further analysis.The GC content estimates the amount of the guanine-cytosine in the nucleotide sequences. The GC3 infers the frequency of either G orC nucleotides present in the third position of the synonymous codon. This does not apply to methionine, tryptophan and the termination codon. These values have a direct correlation with Nc. It measures the synonymous codon usage of genes and its value ranges from 20-62(9).

$$\mathrm{Nc} = 2 + \mathrm{S} + \left(\frac{29}{\mathrm{S}^2 + [1 - \mathrm{S}^2]}\right)$$

The frequency of optimal codons*(Fop)* is the fraction of synonymous codons that are optimally used. It is given by (*Fop) =Noc/ Nsc* where N represents the frequency of each codon, *Noc* and *Nsc* represent optimal codons and synonymous codons respectively. The *Fop* values range from 0 to 1. If the value of *Fop* is 1, it shows the usage of all optimal codons.

GRAVY scores determine the hydropathic indices of amino acids (Peden, 1999). A positive score indicates the hydrophobic nature and a negative score shows the hydrophilic nature of aminoacids.

RSCU-Relative synonymous codon usage

The Relative Synonymous Codon Usage (RSCU) values for the genes are calculated to understand the characteristics of synonymous codon usage without the confounding influence of the amino acid composition of different gene samples (9). The codons with RSCU values >1.0 have positive codon usage bias (abundant codons), while those with RSCU values <1.0 have negative codon usage bias (less-abundant codons); and when the RSCU values are 1.0, it means that these codons are chosen equally or randomly, indicates lack of bias (9). The RSCU is the observed frequency of a codon divided by the frequency expected if all synonymous codons for that amino acidare used equally. The synonymous codons with RSCU more than 1.6 are thought to be over-represented (Fig 1), while the synonymous codons with RSCU less than 0.6 are regarded as under-represented (9). The RSCU values are particularly useful in comparing codon usage between genes that differ in size and amino acid composition.

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	1478	0.15	**Ser**	UCU	1478	0.26
	UUC	**18,420**	1.85		**UCC**	**9110**	**1.59**
Leu	UUA	439	0.05		UCA	2445	0.43
	UUG	6377	0.76		**UCG**	**11,497**	**2.01**
	CUU	6380	0.52		AGU	1944	0.34
	CUC	**19,480**	**2.31**		**AGC**	**7792**	**1.36**
	CUA	897	0.11	**Cys**	UGU	1176	0.36
	CUG	**19,096**	**2.26**		**UGC**	**5387**	**1.64**
Tyr	UAU	5356	0.70	**Pro**	CCU	2239	0.33
	UAC	10,055	1.30		CCC	6481	0.95
His	CAU	7382	0.99		CCA	3221	0.47
	CAC	**7523**	**1.01**		**CCG**	**15,453**	**2.26**
Glu	CAA	3912	0.37	**Arg**	**CGU**	**7844**	**1.24**
	CAG	**17,336**	**1.63**		**CGC**	**18,196**	**2.87**
Ile	AUU	7265	0.71		CGA	2919	0.46
	AUC	21,263	2.07		CGG	5063	0.80
	AUA	2230	0.22		AGA	1534	0.24
Asn	AAU	6393	0.64		AGG	2415	0.38
	AAC	13,665	1.36	**Thr**	ACU	2018	0.24
Lys	AAA	4807	0.47		**ACC**	**15,580**	**1.84**
	AAG	15,864	1.53		ACA	3548	0.42
Val	GUU	3108	0.27		**ACG**	**12,781**	**1.51**
	GUC	**13,246**	**1.17**	**Ala**	GCU	3800	0.25
	GUA	2313	0.20		**GCC**	**26,109**	**1.71**
	GUG	**26,589**	**2.35**		GCA	10,012	0.65
Asp	GAU	**13,045**	**0.71**		**GCG**	**21,326**	**1.39**
	GAC	**23,821**	**1.29**	**Gly**	GGU	7659	0.68
Glu	GAA	12,723	0.73		**GGC**	**26,407**	**2.35**
	GAG	**22,073**	**1.27**		GGA	5299	0.47
					GGG	5499	0.49

Fig 1: A representation of RSCU table

The CODON W software is used to calculate the correspondence analysis of codon count and amino acid frequencies. The file containing gene sequences is loaded in CODON W. For calculating the correspondence analysis option 5 is selected. Run the program and the output 'genes.coa' is selected for downstream processing.

$$\text{RSCU} = \frac{\text{Frequency of codon}}{\text{Expected frequency of codon (if codon usage is uniform)}}$$

CAI-Codon Adaptation Index

Codon adaptation index is a widely used index for studying gene expression in general and efficiency of translation in particular.CAI has been used extensively in biological research. It has been used to study functional conservation of gene expression across different microbial species (10), to predict protein production (11, 12), and to optimize DNA vaccines (13). CAI has recently been used for detecting lateral gene transfer (14).

The cai program in EMBOSS (15), typically referred to as the EMBOSS.cai program is most popularly used. Software for computing CAI is a web application called CAI Calculator 2 (16). The improved CAI is implemented as a new function in DAMBE (17, freely available at http://dambe.bio.uottawa.ca/dambe.asp), which uses a windowed user interface. DAMBE can read 20 standard sequence file-formats including files in the simple FASTA format and the more involved GenBank format or trace files from automatic sequencers. The CAI function can be accessed by clicking 'Seq. Analysis|Codonusage|CAI'. The ensuing dialog box is self-explanatory, except that, for species without a reference set of highly expressed genes, a codon table based on tRNA anticodon can be used by clicking the alternative option button. CAI values vary from 0 to 1 and higher CAI values indicate that the gene of interest has a codon usage pattern more similar to the highly expressed genes.

Codon usage bias (CUB) is usually defined as a species-specific deviation from uniform codon usage in the coding regions of genomic sequences. This bias is possible due to the redundancy of the genetic code, which allows differential use of synonymous codons.The particular pattern of bias observed in a given species is thought to be the product of drift and selection pressures acting on several parameters, but mainly on tRNA gene copy number and genomic %GC content. CUB is therefore a strong species-specific statistic with numerous applications, such as gene prediction or the identification of laterally transferred genes.

Statistical analysis

All the statistical analysis can be carried out using IBM SPSS Statistics 21 software.The Pearson and the Spearman rank correlation coefficients and their corresponding *P*-values can be computed integrally using Microsoft Excel. The Pearson correlations may be computed using built-in Excel functions. The Spearman rank correlations can be computed using the statistical function set of WimGielis. Throughout the manuscript, the Pearson correlation coefficients are denoted by *r* and the Spearman rank correlation coefficients by ρ. The asterisk rating system is used for correlation *P*-values [single asterisk (*), $P < 0.05$–0.01; double asterisks (**), $P < 0.01$–0.001; triple asterisks (***), $P < 0.001$]. *P*-values are relative to a two-tailed Student's *t*-test on the null hypothesis (no correlation). For multiple correlations, we report the mean and standard deviation of the correlation coefficient and the largest *P*-value among the correlations.

Protein energetic cost

It can be defined as the energy consumed for the synthesis of an amino acid encoded by a specific functional codon.Mostly, the energy cost of potentially highly expressed genes is lower than the energy budget of the rest of the proteome. But it cannot be applied to all organisms. In the case of actinobacteria, it has been shown that energy cost varies with its niche (18)

CAI is a major index to measure the m RNA expression level.Generally,genes with10% of the highest and lowest CAI value are chosen as potentially highly expressed and potentially lowly expressed genes respectively. The remaining genes are considered as potentially medially expressed genes.DAMBE software calculates the EC(Dambe ver.6.4.81).The EC values can be analyzed statistically using ANOVA test,F-test and t-test. Heat maps are generated using R statistical software.

tRNA Adaptation index

Estimation of t RNA usage by the coding sequences of a genome is termed as t RNA Adaptation index (tAI). It measures the availability of t RNAsfor every codon of a coding sequence and estimates the level of adaptation between a coding sequence and estimates the level of adaptation between a coding sequenceand the corresponding tRNA pool of the cell. tAI is estimated using the codonR scripts downloaded from http://people.cryst.bbk.ac.uk/*fdosr01/tAI/. it uses the formula

$$tAI_g = \left(\prod_{k=1}^{l_g} w_{ikg}\right)$$

l_g=the length the gene in codons; w_{ikg}= relative adaptability of the codon defined by the k-th triplet in the gene.

Blast Matrix

Blast matrix (BM) is the visual representation of the pair-wise alignment of sequences using the BLAST algorithm with 50/50 rule (5) (Fig 2).The genomic and proteomic sequences of the genomes are downloaded from IMG database. Coding sequences with 300 nucleotides or more and with proper initiation and termination codons are taken for this analysis. If 50% of the alignment is identical with the

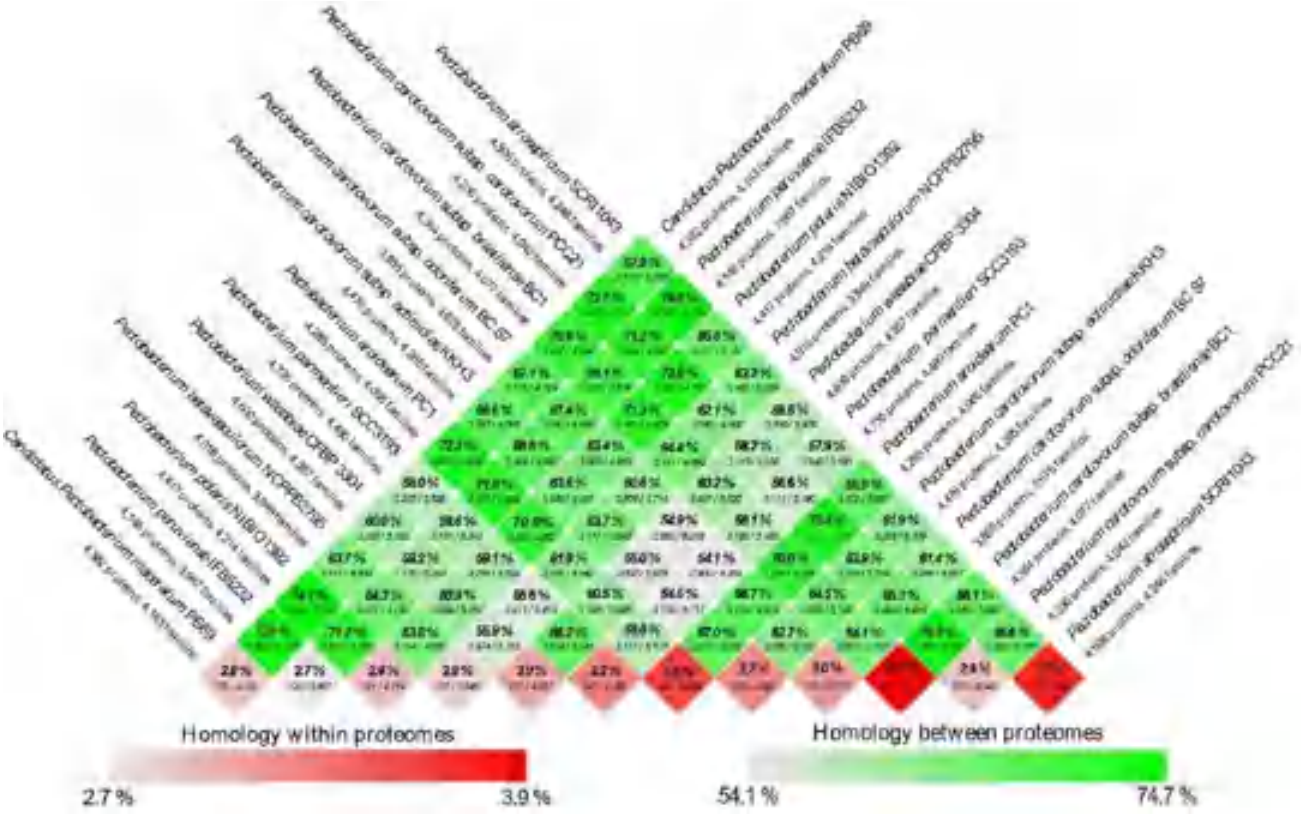

Fig 2: A representation of Blast Matrix

longest protein in the comparative study the BLAST hit is significant. The two sequences are assigned to a “protein family” if they share a similar cut-off value. The amount of shared proteins between the two proteomes is indicated as shaded green in the blast result. In BM, the homolog proteins within the proteome are shown in shaded redcolor at the bottom of the matrix. The color scales are automatically set from highest to lowest value observed. Two programs namely “matrix_createconfig”and “matrix” implemented in CMG Biotools software are used for creating BM (5).

Pan-Core genome plot

Pan-genome of a bacterial species refers to the gene families of all concerned strains of interest, representing a particular species (5) (Fig. 3). The core genome is defined as the conserved pool of genes shared between all strains of a particular species. The homology of genes within and across the genomes is identified based on sequence similarity. The genes are translated into amino acid sequences and aligned all against all using BLASTP. The two genes are considered as gene pair if the alignment follows '50/50' rule,i.e.; if the amino acid sequences are more than 50%identical over more than the length of their length.The common genes of all gene pairs are compiled into a gene family of a genome.

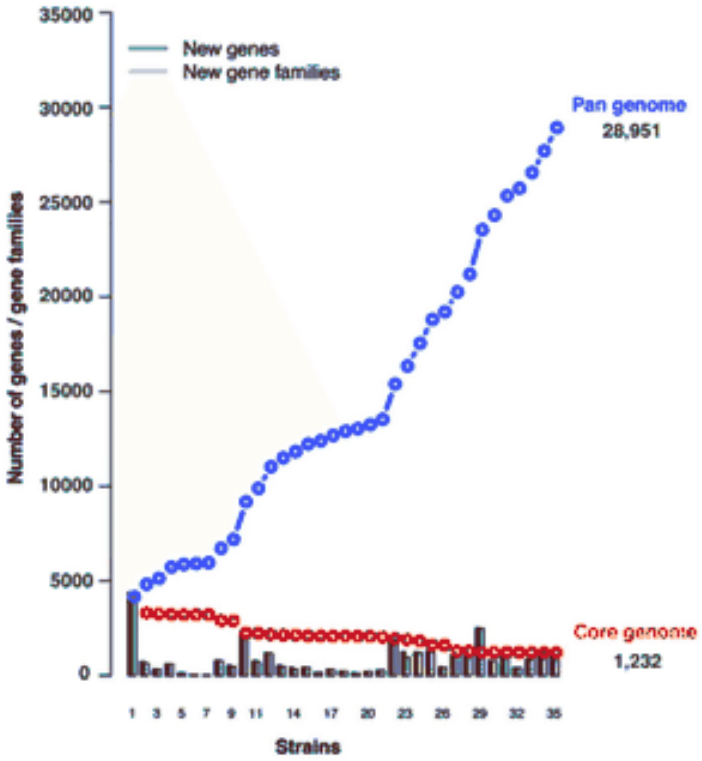

Fig 3: A representation of pan core plot analysis

Two programs namely, "pancoreplot_createConfig"and "pancoreplot" present in CMG Biotool software is adopted for creating the pan-core genome plot and identifying the core genomes among the strains under study.

Metagenomics for the uncultivable microbes

Uncultured microorganisms are mostly responsible for natural biodiversity on Earth. Typically, more than 99% of microorganisms from natural ecosystems are uncultured under laboratory conditions. Therefore,

there is a demand for “culture-independent” approaches for identification and characterization of such microorganisms to uncover their ecological roles in the biosphere. Metagenomics is culture-independent, sequencing-based and/or function-based analysis of the collective genome of a microbial community, enabling us for collection of essential information about community structure and genetic and metabolic potential of the members. This provides insights into the biology of these microorganisms. Another culture-independent method is single-cell genomics, which obtains information about microbial population based on the isolation and genome sequencing of a single cell.

Phylogeny and evolutionary analysis

Phylogenetic tree can be generated based on 16S, 23S, concatenated housekeeping genes and whole genomes(1)(Fig. 4). Most used algorithm in the field of Bioinformatics is Neighbor-joining phylogenetic tree (based on the Nucleotide: Maximum Composite Likelihood method) generated using Mega6.0 software (Tamura et al. 2013). Recently we have developed another method for generating phylogenetic tree and that is binary tree based on the present and absent of phylogeny. The presence and absence of a domain can be designated by 1 and 0 respectively. NTsys software may be used to build the phylogenetic tree (Sarkar et al. 2019).

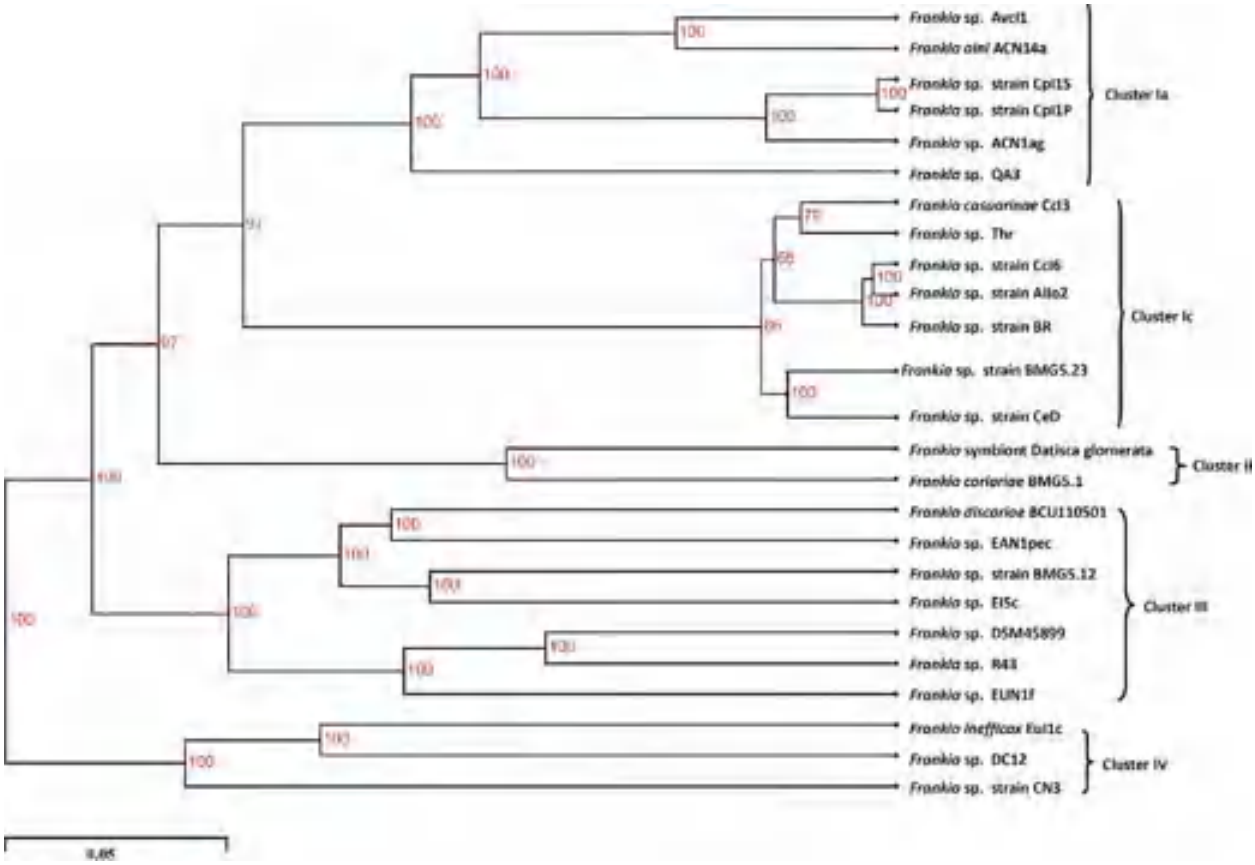

Fig 4: A representation of MLSA based phylogeny of *Frankia*

Kimura two parameter (K2P) model can be employed to generate a distance matrix based on the relative distances of the various clades that resulted from the phylogenetic tree and the corresponding K2P distances can beinvestigated for statistical analysis. The clades (groups) generated from the phylogenetic tree can be treated separately and the outgroup member in every clade should be considered as the reference strain against which the orthologous genes have to be screened in the other strains of that particular clade. Orthologous sequences can be identified using Reciprocal Best Blast Hit approach keeping an identity level of C50 %, an E-value of 1e-10 and at least 50 % region of alignment, using the local BLASTP program (http://www.ncbi.nlm.nlm.nih.gov/BLAST/down load.shtml).

The ratio (ω) of rate of non-synonymous substitutions per non-synonymous site (Ka) to rate of synonymous substitutions per synonymous site (Ks) has been an excellent estimator of the evolutionary selection constraint on a protein-coding gene. ω<1 signifies positive (diversifying) Darwinian selection whereas ω>1 symbolizes purifying (refining) selection. At neutral evolutionary stage, ω = 1, i.e., the rate of synonymous and non-synonymous substitutions are equal. The evolutionary rates of the orthologous protein coding genes can be calculated using Codeml program in the PAML software package (ver. 4.5) (http://abacus.gene.ucl.ac.uk/software/paml.html) with runmode = -2 and CodonFreq = 1. A BioPerl script, developed by us, can also be used along with the Codeml package that translated the cDNAs into proteins and aligned them accordingly. The protein alignments are then projected back into cDNA coordinates that should be used by the PAML package to perform the evolutionary rate analysis employing maximum likelihood method.

Our group	Target genome	dn/ds				
		PHX	S	CAZyme	S	PLX
Frankia alni ACN14a (C-Ia)	*Frankia* sp. AvcL1	0.19	>	0.14	<	0.26
Frankia sp. QA3 (C-Ia)	*Frankia* sp. ACN1aq	0.22	>	0.12	<	0.29
	Frankia sp. CpI1-P	0.24	>	0.11	<	0.35
	Frankia sp. CpI1-S	0.20	>	0.16	<	0.34
Frankia sp. Thr (C-Ic)	*Frankia casuarinae* Ccl3	0.20	>	0.15	<	0.23
Frankia sp. BR (C-Ic)	*Frankia* sp. Allo2	0.24	>	0.15	<	0.26
	Frankia sp. Ccl6	0.18	>	0.09	<	0.22
Frankia sp. CeD (C-Ic)	*Frankia* sp. BMG5.23	0.22	>	0.12	<	0.24
Frankia coriariae BMG5.1 (C-II)	*Frankia* Datisca glomerata Dg1	0.13	>	0.05	<	0.17
Frankia sp. EAN1pec (C-III)	*Frankia discariae* BCU110501	0.15	>	0.09	<	0.22
Frankia sp. EI5c (C-III)	*Frankia* sp. BMG5.12	0.14	>	0.05	<	0.19
Frankia sp. EUN1f (C-III)	*Frankia* sp. DSM 45899	0.14	>	0.07	<	0.22
	Frankia sp. R43	0.17	>	0.09	<	0.23
Frankia sp. CN3 (C-IV)	*Frankia* sp. DC12	0.18	N	0.17	<	0.24
	Frankia inefficax EuIIc	0.16	N	0.15	<	0.22

S Statistically significant ($p < 0.001$)

N Not statistically significant

Fig 4: A representative table for evolutionary analysis based on dN/dS score

Pairs of sequences with Ka and Ks values, evocative of saturation, should not be considered for further analysis. Evolutionary selection pressure among the differentially expressed genes can also be estimated to decipher the varying tendencies of evolution among studied genes. Evolutionary rates of the genes transcribed from the complimentary strands of replication (leading and lagging strands) can also be assessed separately for all the genomes under analysis.

Conclusion

In this book chapter, we have discussed some of the popular bioinformatics techniques used for downstream analysis of post-genomic era. All these techniques if, used in appropriate way will lead researchers to statistical results with proper visualization of data. Codon Usage, Amino acid usage, energy cost analysis along with phylogeny and evolutionary analysis will gather knowledge on culturable strains whose whole genome sequence is available. On contrary, 16s metagenomics will help us in identifying the unculturable micro-organisms present in specific niche. Thus, through combination of all these approaches, a holistic idea about Actinobacterial life style can be studied.

REFERENCES:

1. Sen, A., V. Daubin, D. Abrouk, I. Gifford, A.M. Berry, and P. Normand. 2014. Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders 'Frankiales' and Micrococcales should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov.andNakamurellales ord. nov.Int J SystEvolMicrobiol 64:3821-32.

2.Ventura, M., C. Canchaya, A. Tauch, G. Chandra, G.F. Fitzgerald, K.F. Chater, and D. van Sinderen. 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. MicrobiolMolBiol Rev 71:495-548

3. ShuvankarBallav , Syed G Dastager and Savita Kerkar 2012. Biotechnological significance of Actinobacterial research in India. Recent Research in Science and Technology 2012, 4(4): 31-39

4. Aderem, A. 2005. Systems biology: its practice and challenges. Cell 121:511-3.

5. Bork, Peer & Serrano, Luis. (2005). Towards Cellular Systems in 4D. Cell. 121. 507-9. 10.1016/j.cell.2005.05.001.

6. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 1985;2: 13–34.

7. Sharp and Li (1986) Codon usage in regulatory genes in Escherichia coli does not reflect selection for 'rare' codons. Nucleic Acids Res1986 Oct 10;14(19):7737-49

8. Wright .F (1990) The 'effective number of codons' used in a gene. Gene. 87 (1990) 23-29

9. Peden JF (1999) Codon W. PhD Dissertation, University of Nottingham, Nottinghamshire, UK

10. Lithwick, Gila &Margalit, Hanah. (2005). Relative predicted protein levels of functionally associated proteins are conserved across organisms. Nucleic acids research. 33. 1051-7. 10.1093/nar/gki261.

11. Futcher B, Latter GI, Monardo P, McLaughlin CS, Garrels JI. A sampling of the yeast proteome. Mol Cell Biol. 1999; 19(11):7357-7368. doi:10.1128/mcb.19.11.7357

12. Grosjean, H., &Fiers, W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene, 18(3), 199-209.

13. Ruiz LM, Armengol G, Habeych E, Orduz S. A theoretical analysis of codon adaptation index of the Boophilusmicroplus bm86 gene directed to the optimization of a DNA vaccine. J. Theor. Biol. 2006;239:445–9.

14. Bodilis J, Barray S. Molecular evolution of the major outer-membrane protein gene (oprF) of Pseudomonas. Microbiology. 2006;152:1075–88

15. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16:276–7

16. Wu G, Culley DE, Zhang W. Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. Microbiology. 2005;151:2175–87.

17. Xia X, Xie Z. DAMBE: Software package for data analysis in molecular biology and evolution. Journal of Heredity. 2001;92:371–373.

18. Sarkar, I., Gtari, M., Tisa, L. S., & Sen, A. (2019). A novel phylogenetic tree based on the presence of protein domains in selected actinobacteria. Antonie van Leeuwenhoek, 112(1), 101-107.

19. Emily HM Wong1 , David K Smith2*, Raul Rabadan3 , Malik Peiris1 , Leo LM Poon1*Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. BMC Evolutionary Biology 2010, 10:253

20. ErmolaevaM.D.Synonymous codon usage in bacteria. Curr. Issues Mol. Biol. 2001; 3:91–7.

21.Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19(3):1720-1730. doi:10.1128/mcb.19.3.1720

22. Tsai, Chih-Tung & Lin, Chih-Hung & Chang, Chi-Yao. (2007). Analysis of codon usage bias and base compositional constraints in iridovirus genomes. Virus research. 126. 196-206. 10.1016/j.virusres.2007.03.001.