Review Article

Suitability and advantages of abundance-weighted algorithms for the holistic comparison of sample systems

Abstract

Numerous microcosmic research techniques have been available for examinations of individual chemicals (DNA, RNA, proteins, carbohydrates and other small-molecule chemicals) in sample systems. However, the results generated from many of the microcosmic studies have fueled speculations, hypotheses and non-conclusive debates in the study of systematics, similar to blind men each touching a portion of an elephant. To address this shortcoming, macrocosmic holistic techniques exhibit advantages in profiling and systematically comparing genetic diversity and genetic distance between sample systems, transcriptome and proteome expressions and metabolome/chemical constituent fingerprints. Although binary abundance-unweighted algorithms have been widely used in such holistic comparisons, most of the research data do not fall into an "all-or-none" qualitative category, leading to unsuitability of abundance-unweighted algorithms as analytic tools in these types of holistic comparisons. Using an RAPD molecular polymorphism study as the example, this paper reviews the prerequisites and limitations of the binary abundance-unweighted algorithms in holistic comparisons and the advantage of using abundance-weighted algorithms, which are mathematically general and suitable for any holistic analyses.

Keywords: abundance-weighted similarity computation, abundance-weighted cluster construction; polymorphism of molecular markers, holistic similarity, phylogenetics analysis, natural *cordyceps sinensis*

Lu-Qun Ni,[1] Jia-Shi Zhu,[2,3] Yi-Sang Yao,[2] Yu-Ling Li,[4,5] Wei-Dong Xie[2]

[1]Department of Mathematics, University of California San Diego, USA
[2]Department of Applied Biology and Chemistry Technology, The Hong Kong Polytechnic University, Hong Kong
[3]Division of Life Sciences and Health, Tsinghua University Graduate School at Shenzhen, China
[4]Qinghai Academy of Animal Husbandry and Veterinary Sciences, Qinghai University, China
[5]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, China

Correspondence: Jia-Shi Zhu, Hong Kong Polytechnic University, Hong Kong, Tel +1(858)705-3790, Email zhujosh@gmail.com

Abbreviations: ITS, internal transcribed spacer; CAPS, cleaved amplified polymorphic sequence; DAF, DNA amplified fingerprints; ISSR, inter-simple sequence repeat; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; SCAR, sequence characterized amplified regions; SSCP, single-strand conformation polymorphism; SSR, simple sequence repeat; HS, hirsutella sinensis

Introduction

Although numerous microcosmic research techniques have been developed and widely used to examine individual chemicals (DNA, RNA, proteins, carbohydrates and other small-molecule chemicals) in sample systems, macrocosmic holistic techniques exhibit advantages in systematic comparisons of genetic diversity and genetic distance between sample systems, transcriptome and proteome expressions and metabolome/chemical constituent fingerprints. In holistic research projects, binary abundance-unweighted algorithms have been widely used and are limited to two sample comparisons. However, most of data generated from holistic studies do not fall into an "all-or-none" qualitative research category; thus, abundance-unweighted algorithms may not serve as suitable analytic tools for these types of research. Using an RAPD molecular polymorphism study as the example, this paper reviews the prerequisites and limitations of using the binary abundance-unweighted algorithms and the advantage of abundance-weighted algorithms in holistic studies.

Since 1999, PCR-based nrDNA Internal Transcribed Spacer (ITS) sequencing has been used to determine the taxonomic status of the examined fungal specimens in the natural *Cordyceps sinensis* insect-fungi complex as an auxiliary molecular method to examine the anamorph-teleomorph connection of *Ophiocordyceps sinensis*.[1–4] Scientists have reported the molecular heterogeneity of *C. sinensis*-associated fungi of the genera *Hirsutella, Paecilomyces* and *Tolypocladium* as well as *Geomyces pannorum, Cladosporium macrocarpum, Phaeosphaeria pontiformis* and *Neosetophoma samarorum* (in total, more than 90 species spanning at least 37 genera).[5–21] SNP mass spec genotyping and amplicon sequencing with or without molecular cloning have identified the mutant ITS sequences of at least 17 genotypes of *O. sinensis* fungi with multiple, scattered point mutations or DNA segment substitution hereditary variations from the natural *C. sinensis* insect-fungi complex.[1–4,10–13,19,22–28] The successful production of artificial *C. sinensis* enabled the molecular examination of the fungal strains that were used as inoculation agents and the correlation of the genotype results with the finished product, artificial *C. sinensis*.[28] They reported the identification of the only fungal genotype, Genotype #1 *H. sinensis*, from 3 anamorphic fungal strains that were used as the inoculation agents and of the sole teleomorphic Genotype #4 of *O. sinensis* in the fruiting body and caterpillar body of artificial *C. sinensis*. Therefore, this paper presented the paradox of "planting melon seeds and harvesting beans" and provides strong evidence against the sole *O. sinensis* anamorph hypothesis for Genotype #1 *H. sinensis*, which was proposed a decade ago by the same corresponding author and colleagues.[29]

The microcosmic mycological and molecular results have fueled speculations, paradoxes, hypotheses and non-conclusive debates in mycological systematics studies, similar to blind men each touching a portion of an elephant. To address this shortcoming, the macrocosmic molecular marker polymorphism approach has been

used as a component of overall molecular systematics strategies to profile the natural *C. sinensis* insect-fungi complex as a holistic entity and to compare the holistic polymorphic similarities of the systems without the requirement of precise examinations of the DNA sequences or the individual taxonomies of the component fungi. These macrocosmic molecular techniques for holistic comparisons include AFLP (Amplified Fragment Length Polymorphism), CAPS (Cleaved Amplified Polymorphic Sequence), DAF (DNA Amplified Fingerprints), ISSR (Inter-Simple Sequence Repeat), RAPD (Random Amplified Polymorphic DNA), RFLP (Restriction Fragment Length Polymorphism), SCAR (Sequence Characterized Amplified Regions), SSCP (single-strand conformation polymorphism) and SSR (Simple Sequence Repeat).[1–4,30,31] Among these methodologies, RAPD molecular marker polymorphism analysis is the most frequently used technique for comparing overall similarities or dissimilarities (genetic distances) and exploring the phylogenetic (cluster) relationship between the test systems,[2,29,32–41] although it has been suggested that ISSR may be more sensitive than RAPD[42–44] and costly metagenomic approaches may demonstrate advantages in qualitative studies of microbial genetic diversity and molecular ecology.[45–47] Among the key factors in study design for molecular marker polymorphism comparisons, the importance of unbiased selection of a plurality of random primers for random amplification of the genomic DNA templates isolated from the examined systems has been addressed.[1–4,40,41] The use of only a few random primers without reporting the objectivity and representativeness of the selection could lead to bias in the data analysis and thus biased conclusions.[1–4,40,41,48] When *C. sinensis* insect-fungi complex samples and fungal strain samples were not profiled as a whole, this resulted in the inaccuracy of holistic comparisons, interpretations and conclusions.[29] Because holistic analysis of molecular marker polymorphisms relies upon similarity computation and phylogenetic (cluster) tree construction, selection of analytical algorithms becomes the next question in study design for precise analyses and comparisons of the holistic polymorphisms between study systems.

Computational biology algorithms for polymorphism similarity analysis

Numerous studies have consistently used the PCR amplicon abundance-unweighted algorithm known as the Nei-Li Coefficient equation (1)[49] or similar[50] for similarity computations. The Nei-Li equation:

$$S = \frac{2N_{AB}}{N_A + N_B} \qquad (1)$$

Where

S=Similarity

N_{AB}=Number of matched DNA bands in lanes A and B amplified through PCR and

N_A and N_B=Total number of DNA bands in lanes A and B, respectively.

This equation was designed as an abundance-unweighted algorithm to analyze the "all or none" data, to compare pure systems in pairs and, in particular, to analyze the loss of restriction sites due to mutations.[49] The proper use of this and similar algorithms has 2 computational biological prerequisites:

i. All matched DNA pairs in the electrophoretic lanes being compared must have essentially the same densities and

ii. All DNA amplicons must be well separated from the adjacent DNA moieties with similar molecular weights and conformations by electrophoresis.[40,41]

However, in reality, as shown in Figure 1, many matches have similar or dissimilar densities for the DNA bands of the pairs of electrophoretic lanes and some amplicons are not well separated from the adjacent DNA moieties. Analysis using the Nei-Li equation (1)[49] neglected the proper weights of those high-density amplicons in the similarity computation, such as the 570-, 1,700-, 1,974-bp bands in the left panel of Figure 1 and the 973-bp band in the right panel. This resulted in the multiple low-density amplicon bands being incorrectly over-weighted in the similarity computation and those high-density amplicon bands being largely relatively devalued. Table 1 compares the similarities computed with use of the abundance-unweighted Nei-Li equation (1) and the abundance-weighted ZUNIX equation (2). DNA amplicons were amplified by PCR using random primers S31 (Left Panel of Figure 1) or OPB-01 (Right Panel). The genomic DNA used as the template for PCR was isolated from the ascocarp portion (AC) of wild *C. sinensis* and *Hirsutella sinensis* (Hs) mycelia.
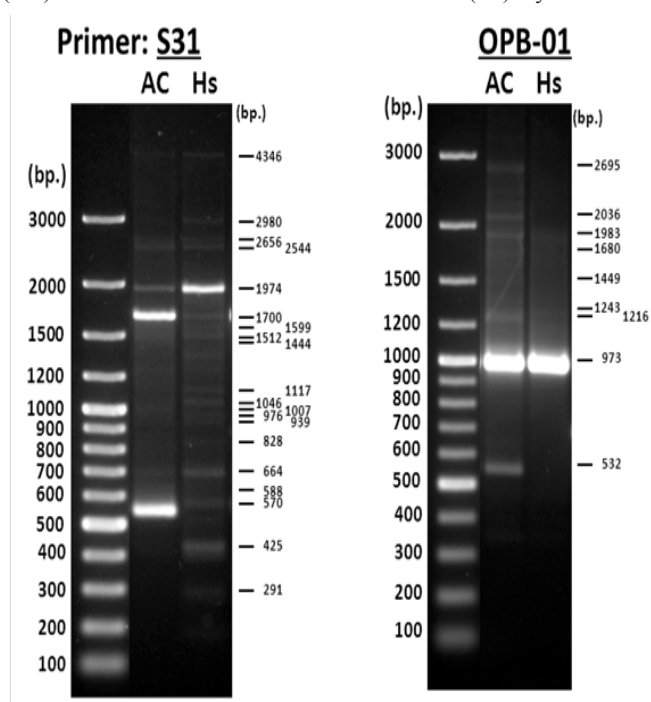

Figure 1 Agarose gel electrophoresis examining RAPD molecular marker polymorphisms of the ascocarp portion (AC) of wild *Cordyceps sinensis* and *Hirsutella sinensis* (HS) with using primers S31 and OPB-01 (Figure 3 of the original paper[41] with re-use permission from Beijing Da Xue Xue Bao Yixueban).

Due to non-satisfaction of the aforementioned computational biology prerequisites, the Nei et al.[49] equation (1) is clearly not suitable for similarity computations for the RAPD studies of *C. sinensis* and results in a loss of molecular information provided by the abundance/density of the DNA amplicons during RAPD polymorphism analysis because of its primitive understanding of reality. New arithmetic methods for similarity computation are required to quantify the molecular information provided by the abundance/density of the

232

DNA amplicons, which are lost or partially lost during RAPD polymorphism analysis using the Nei-Li equation (1).[49] Accordingly, the ZUNIX equations were formulated (The software for ZUNIX algorithms can be downloaded for free from www.ebioland.com/ZUNIX.htm; Beijing Bioland Technology, 2013[40]) for abundance-weighted similarity computation. These new arithmetic methods for similarity computation are required to quantify the molecular information provided by the abundance/density and migration speed of the DNA amplicons in gel electrophoresis. The ZUNIX equation (2) was formulated to compare the polymorphisms of 2 electrophoretic lanes, where $d_{ik} \geq 0$, $i=1,2$ and $k=1,2, \ldots, m$ define the measure of similarity, as follows:[40]

$$S = \frac{\sum\limits_{k=1}^{m} \left[2\, Min\{d_{1k}, d_{2k}\} \right]}{\sum\limits_{k=1}^{m} d_{1k} + d_{2k}} \tag{2}$$

where the similarity of the 2 densities d_{1k} and d_{2k} is the common portion of their values. The ZUNIX equation (2) defines similarity as the total density of all common parts present in the samples of comparison divided by the total density of all bands across the samples. This algorithm is mathematically general with no specific prerequisites and governs all conditions[40] in contrast to the abundance-unweighted Nei-Li equation (1),[49] which narrows the specific cases under the strict prerequisites (a) and (b) described above. The second ZUNIX equation (3) is suitable for comparing the DNA amplicons in 3 electrophoretic lanes, where $d_{ik} \geq 0$, $i=1,2,3$ and $k=1,2, \ldots, m$ and are shown below.[40]

$$S = \frac{\sum\limits_{k=1}^{m} \left[3\, Min\{d_{1k}, d_{2k}, d_{3k}\} \right]}{\sum\limits_{l=1}^{m} d_{1l} + d_{2l} + d_{3l}} \tag{3}$$

Extending further, the third ZUNIX equation (4) is suitable for comparisons of DNA amplicons in more than 3 lanes, where $d_{ik} \geq 0$, $i=1,2, \ldots, n$ and $k=1,2, \ldots, m$ and are described as[40]

$$S = \frac{\sum\limits_{k=1}^{m} \left[n\, Min\{d_{1k}, d_{2k}, \ldots, d_{nk}\} \right]}{\sum\limits_{r=1}^{n} \sum\limits_{s=1}^{m} d_{rs}} \tag{4}$$

Table 1 Comparisons of the polymorphism similarities computed using different algorithms

Primers	Similarity		% Change (un weighted *vs.* weighted)
	By abundance-un weighted Nei-Li Equation (1)	**By abundance-weighted ZUNIX Equation (2)**	
S31	0.5	0.23	-54%
OPB-01	0.29	0.83	+189%

Note: The molecular marker polymorphisms of the ascocarp portion (AC) of *Cordyceps sinensis* and the mycelial culture of *Hirsutella sinensis* (Hs) were examined by PCR amplification using random primers S31 and OPB-01 and by agarose gel electrophoresis, as shown in Figure 1. The overall polymorphism similarities were computed using either the abundance-unweighted Nei-Li equation (1) or the abundance-weighted ZUNIX equation (2).

The ZUNIX equations (2) (3) (4) arithmetically consider the following:

i. The unmatched DNA (or protein or other chemical) bands and their densities (abundance)

ii. Differences in the density (abundance) of the matched DNA (or protein or other chemical) bands (or peaks or areas under the curves) and

iii. The ability to compare multiple samples.

The abundance-weighted ZUNIX equations (2) (3) (4) define similarity as the total density (abundance) of all common parts present in the matched DNA bands of the samples being compared divided by the total density (abundance) of all bands across the samples.[40] The ZUNIX equations are mathematically general, with no specific prerequisites, and govern all conditions including the special cases under the strict prerequisites set forth by the Nei-Li equation (1).[49] The ZUNIX equations accurately capture all of the molecular information buried in the amplicon DNA bands (both the abundance/density and the migration speed in gel electrophoresis) in the RAPD (or ISSR, SSCP, or other holistic analysis techniques) gel images, which are partially lost or even significantly lost when incorrectly using the abundance-unweighted Nei-Li equation (1).[49] Consequently, misuse of the Nei-Li equation (1)[49] in *C. sinensis* holistic polymorphism studies when the sample systems do not meet the specific computational biology prerequisites may lead to inaccurate calculations of overall similarities and questionable conclusions.[35,44] Yao et al.[41] employed abundance-weighted ZUNIX equations (2) (3) (4) to analyze polymorphic data (1,418 amplicon bands) obtained from 20 RAPD gel images using 20 random primers and reported similarities of 0.546-0.686 between *H. sinensis* and the stroma, caterpillar body and ascocarps of natural *C. sinensis*. The results are inconsistent with the sole *O. sinensis* anamorph hypothesis for Genotype #1 *H. sinensis*, which was proposed by Wei et al.[29] and Guo et al.[51] The mathematically general, abundance-weighted ZUNIX equations can also accurately calculate the similarities of mass spec proteomic polymorphisms of multiple *C. sinensis* samples and the similarities of HPLC fingerprints and other chemical chromatographic and electrophoretic profiles.[40,52]

Abundance-weighted algorithms for phylogenetic tree construction

In previous RAPD and ISSR studies of *C. sinensis*, phylogenetic analysis primarily used binary PCR amplicon abundance-unweighted UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm to construct phylogenetic trees.[29,33,34,38,43,44] Consistent with the computation of similarity, constructing phylogenetic trees using abundance-unweighted algorithms in the holistic analysis negates the differences between high- and low-density DNA amplicon bands or between the complete and incomplete separation of DNA amplicon moieties by agarose gel electrophoresis and their impact

on the weights in exploring phylogenetic similarity and dissimilarity, leading to computational errors in constructing phylogenetic trees. Ni et al.[40] demonstrated the inaccurate construction of a phylogenetic tree improperly using the abundance-unweighted algorithm, reflecting the inability to capture and analyze all molecular information buried in the DNA bands (both the abundance/density and the migration speed in agarose gel electrophoresis) in RAPD gel images, whereas the abundance-weighted algorithms corrected such analytical errors. Thus, the selection of different clustering algorithms with or without consideration of the abundance/densities and incomplete separation of the DNA moieties (or RNAs, proteins, or other chemicals) greatly impacts the data analysis and study conclusions and the abundance-unweighted algorithm is not suitable for studies of *C. sinensis*, which contains multiple intrinsic fungi.[40,41]

As indicated by Ni et al.,[40] PAUP 4.0B requires semi-quantitative scoring prior to phylogenetic tree construction when using the abundance-weighted algorithm to construct phylogenetic trees, which may slightly reduce the sensitivity in handling fully quantitative abundance/density data. Therefore, Ni et al.[40] performed their clustering analysis using software with full quantitative capacity, such as Cluster3.0, JMP9 and SPSS and they demonstrated that the fully quantitative algorithms placed *H. sinensis* in a separate clade from the main *C. sinensis* cluster at a large rescaled distance. Both the semi-quantitative and fully quantitative methods, however, exhibited advantages in capturing all molecular information and accurately constructing phylogenetic trees in the *C. sinensis* molecular and proteomic polymorphism studies.[40,41,52] Other advantages of software for fully quantitative clustering include ease of use and accurate quantitation, but the full-quantitative algorithms provided by the aforementioned software do not include bootstrap value calculation, whereas the semi-quantitative clustering algorithm provided by PAUP 4.0B does calculate the bootstrap value (usually Bootstrap=1000). Yao et al.[41] employed the abundance-weighted algorithms to analyze 1,418 DNA amplicon bands obtained from 20 RAPD gel images using 20 random primers and constructed a phylogenetic tree showing an isolated *H. sinensis* leaf that is greatly distant from the *C. sinensis* clade. This large phylogenetic distance and dissimilarities (0.546-0.686) between *H. sinensis* and the stroma, caterpillar body and ascocarp samples of *C. sinensis* determined using the abundance-weighted algorithms is inconsistent with the sole *O. sinensis* anamorph hypothesis for *H. sinensis*. The types of algorithms and software should be carefully considered when designing RAPD, ISSR, SSCP and other holistic profile comparison studies, although both fully and semi-quantitative abundance-weighted clustering algorithms can be used in general for *C. sinensis* molecular and proteomic polymorphism studies.[40,41,52]

The mathematically general, abundance-weighted algorithms have also been employed in a proteomatic polymorphism study,[52] in which the proteins of *C. sinensis* insect-fungi complex and of mycelial cultures of *H. sinensis* and *Paecilomyces hepiali* were extracted and separated through SELDI-TOP mass spectrometry. In this study, Dong et al.[52] analyzed more than 1,900 protein bands, computed the holistic similarities between the samples and constructed a phylogenetic tree. Again, the abundance-weighted algorithms have exhibited advantages in capturing all of the proteomic information buried in the mass spec peaks/bands (both the peak heats or abundance and the migration speed in SELDI-TOP mass spectrometry) and in accurately calculating the similarities of proteomic polymorphisms of the samples. This study reported profound, dynamic, asynchronous changes in holistic proteomic polymorphisms in the stroma and caterpillar body of

natural *C. sinensis* during maturation, reflecting dynamic changes in the expression of transcriptomic genes in the different compartments of *C. sinensis* at different stages of maturation. This study also reported significant differences in holistic proteomic polymorphisms between samples of natural *C. sinensis* and mycelial cultures of *H. sinensis* and *P. hepiali*, indicating that neither of the fungal culture preparations might be a stand-along therapeutic regime to replace natural *C. sinensis*.

Conclusion

These mathematically general, abundance-weighted algorithms can be used with no computational biology prerequisites and limitations for overall similarity computation and phylogenetic (cluster) tree construction in holistic comparison studies of genetic diversity and genetic distance, transcriptome and proteome expressions, metabolomics and chemical constituent fingerprints using techniques including gel electrophoresis, two-dimensional gel electrophoresis, capillary electrophoresis, mass spectrometry, chromatography, etc.

Acknowledgements

Conflict of interest

The author declares no conflict of interest.

References

1. Yao YS, Xie WD, Zhu JS. Genetic heterogeneity of natural *Cordyceps sinensis* with co-existence of multiple fungi and at least 12 genotype mutants. The First National Cordyceps Conference and *Ophiocordyceps sinensis* Industrial Summit; China; 2015. p. 63–82.

2. Zhu JS, Wu JY. Genetic heterogeneity of natural *Cordyceps sinensis* with co-existence of multiple fungi. *Chin J Cell Biol*. 2015;37(2):284–298.

3. Li YL, Yao YS, Xie WD, et al. Molecular heterogeneity of natural *Cordyceps sinensis* with multiple *Ophiocordyceps sinensis* fungi challenges the anamorph-teleomorph connection hypotheses. *Am J Biomed Sci*. 2016;8(2):123–159.

4. Zhu JS, Li YL. *A Precious Transitional Chinese Medicine, Cordyceps sinensis: Multiple heterogeneous Ophiocordyceps sinensis in the insect-fungi complex*. Saarbrüchen, Germany: Lambert Academic Publishing; 2017.

5. Chen YQ, Hu B, Xu F, et al. Genetic variation of *Cordyceps sinensis*, a fruit-body-producing entomopathegenic species from different geographical regions in China. *FEMS Microbiol Lett*. 2004;230(1):153–158.

6. Jiang Y, Yao YJ. ITS sequence analysis and ascoma development of *Pseudogymnoascus roseus*. *Mycotaxon*. 2005;94:55–73.

7. Chen JC, Zhang W, Lu T, et al. Morphological and genetic characterization of a cultivated *Cordyceps sinensis* fungus and its polysaccharide component possessing antioxidant property in H22 tumor-bearing mice. *Life Sci*. 2006;78(23):2742–2748.

8. Leung PH, Zhang QX, Wu JY. Mycelium cultivation, chemical composition and anti-tumour activity of a Tolypocladium sp. fungus isolated from wild *Cordyceps sinensis*. *J Appl Microbiol*. 2006;101(2):275–283.

9. Zhu JS, Guo YL, Yao YS, et al. Maturation of *Cordyceps sinensis* associates with co-existence of *Hirsutella sinensis* and *Paecilomyces hepiali* DNA and dynamic changes in fungal competitive proliferation predominance and chemical profiles. *J Fungal Res*. 2007;5(4):214–224.

10. Zhu JS, Gao L, Li XH, et al. Maturational alterations of oppositely orientated rDNA and differential proliferations of CG:AT-biased genotypes of *Cordyceps sinensis* fungi and *Paecilomyces hepiali* in natural *Cordyceps sinensis*. *Am J Biomed Sci*. 2010;2(3):217–238.

11. Zhu JS, Zhao JG, Gao L, et al. Dynamically altered expressions of at least 6 *Ophiocordyceps sinensis* mutants in the stroma of *Cordyceps sinensis*. *J Fungal Res*. 2012;10(2):100–112.

12. Zhu JS, Yao YS, Li YL, et al. Hereditary variations with DNA segment substitution between fungi in the two types of *Cordyceps sinensis* ascospores suggest fungal chromosomal fusions. The 11th National Conference of Medicinal Mycology, China; 2016. p. 92–103.

13. Yang JL, Xiao W, He HX, et al. Molecular phylogenetic analysis of *Paecilomyces hepiali* and *Cordyceps sinensis*. *Yao Xue Xue Bao*. 2008;43(4):421–426.

14. Zhang YJ, Sun BD, Zhang S, et al. Mycobiotal investigation of natural *Ophiocordyceps sinensis* based on culture-dependent investigation. *Mycosystema*. 2010;29(4):518–527.

15. Zhang Y, Zhang S, Wang M, et al. High diversity of the Fungal community structure in naturally-occurring *Ophiocordyceps sinensis*. *PLoS One*. 2010;5(12):e15570.

16. Barseghyan GS, Holliday JC, Price TC, et al. Growth and cultural-morphological characteristics of vegetative mycelia of medicinal caterpillar fungus *Ophiocordyceps sinensis* GH Sung et al. (Ascomycetes) Isolates from Tibetan Plateau (P.R.China). *Int J Med Mushrooms*. 2011;13(6):565–581.

17. He SQ, Jin XL, Luo JC, et al. Morphological characteristics and submerged culture medium screening of a strain of *Geomyces pannorum* isolated from *Cordyceps sinensis*. *Chin J of Grassland*. 2010;32(z1):70–75.

18. He SQ, Ji XL, Luo JC, et al. Morphological characteristics and submerged culture medium screening of *Pseudogymnoascus roseus*. *Microbiol China*. 2011;38(9):1371–1376.

19. Yao YS, Zhou YJ, Gao L, et al. Dynamic alterations of the differential fungal expressions of *Ophiocordyceps sinensis* and its mutant genotypes in stroma and caterpillar during maturation of natural *Cordyceps sinensis*. *J of Fungal Res*. 2011;9(1):37–53.

20. Xia F, Liu Y, Shen GR, et al. Investigation and analysis of microbiological communities in natural *Ophiocordyceps sinensis*. *Can J Microbiol*. 2015;61(2):104–111.

21. Xia F, Liu Y, Guo MY, et al. Pyrosequencing analysis revealed complex endogenetic microorganism community from natural Dong Chong Xia Cao and its microhabitat. *BMC Microbiol*. 2016;16(1):196.

22. Stensrud Ø, Jones NLH, Schumacher T. Towards a phylogenetic classification of *Cordyceps*: ITS nrDNA sequence data confirm divergent lineages and paraphyly. *Mycol Res*. 2005;109(Pt 1):41–56.

23. Stensrud Ø, Schumacher T, Tabrizi KS, et al. Accelerated nrDNA evolution and profound AT bias in the medicinal fungus *Cordyceps sinensis*. *Mycol Res*. 2007;111(Pt 4):409–415.

24. Xiao W, Yang JL, Zhu P, et al. Non-support of species complex hypothesis of *Cordyceps sinensis* by targeted rDNA-ITS sequence analysis. *Mycosystema*. 2007;28(6):724–730.

25. Gao L, Li XH, Zhao JQ, et al. Detection of multiple *Ophiocordyceps sinensis* mutants in premature stroma of *Cordyceps sinensis* by MassARRAY SNP MALDI-TOF mass spectrum genotyping. *Beijing Daxue Xuebao (Yixueban)*. 2011;43(2):259–266.

26. Gao L, Li XH, Zhao JQ, et al. Maturation of *Cordyceps sinensis* associates with alterations of fungal expressions of multiple *Ophiocordyceps sinensis* mutants with transition and transversion point mutations in stroma of *Cordyceps sinensis*. *Beijing Daxue Xuebao (Yixueban)*. 2012;44(3):454–463.

27. Mao XM, Zhao SM, Cao L, et al. The morphology observation of *Ophiocordyceps sinensis* from different origins. *J Environmental Entomol*. 2013;35(3):343–353.

28. Wei JC, Wei XL, Zheng WF, et al. Species identification and component detection of *Ophiocordyceps sinensis* cultivated by modern industry. *Mycosystema*. 2016;35(4):404–410.

29. Wei XL, Yin XC, Guo YL, et al. Analyses of molecular systematics on *Cordyceps sinensis* and its related taxa. *Mycosystema*. 2006;25(2):192–202.

30. Song DL, Zhang J, Chen JX. DNA molecular markers and their application in *Cordyceps sinensis* research. *Acta Edulis Fungi*. 2002;9(3):52–55.

31. Bai Y. DNA molecular marker techniques and their application. *J Anhui Agri Sci*. 2007;35(24):7422–7424.

32. Chen Y, Wang W, Yang Y, et al. Genetic divergence of *Cordyceps sinensis* as estimated by random amplified polymorphic DNA analysis. *Yi Chuan Xue Bao*. 1997;24(5):410–416.

33. Chen Y, Zhang YP, Yang Y, et al. Genetic diversity and taxonomic implication of *Cordyceps sinensis* as revealed by RAPD markers. *Biochem Genet*. 1999;37(5-6):201–213.

34. Zhang YW, Chen YJ, Shen FR, et al. Study of genetic divergence in *Cordyceps sinensis* and *C. crassispors* from Northwest of Yunnan by using RAPD. *Mycosystema*. 1999;18(2):176–183.

35. Li ZZ, Huang B, Li CR, et al. Molecular evidence for anamorph determination of *Cordyceps sinensis* (Berk.) Sacc. I. Relation between *Hirsutella sinensis* and *C. sinensis*. *Mycosystema*. 2000;19(1):60–64.

36. Yu C, He JL, Zeng W, et al. Fingerprint comparison on *Cordyceps sinensis* and artificial ones by RAPD. *Chin Tradit Herbal Drugs*. 2005;36(2):274–277.

37. Feng K, Wang S, Hu DJ, et al. Random amplified polymorphic DNA (RAPD) analysis and the nucleosides assessment of fungal strains isolated from natural *Cordyceps sinensis*. *J Pharm Biomed Anal*. 2009;50(3):522–526.

38. Qian M, Zeng W, Zhang DL. RAPD polymorphism analysis on *Cordyceps sinensis*. *LisShenzhen Med Materia Medica Res*. 2011;22(7):1738–1739.

39. Xie F, Zhu ZX, Chen JJ. RAPD and analysis of ITS sequences of *Ophiocordyceps sinensis* in Gansu. *Chin J Microecol*. 2011;23(3):219–229.

40. Ni LQ, Yao YS, Gao L, et al. Arithmetic methods for similarity computation and cluster tree construction in RAPD analysis on a multi-fungi microecology system, *Cordyceps sinensis*. *Am J BioMed Sci*. 2014;6(2):82–104.

41. Yao YS, Gao L, Li YL, et al. Amplicon density-weighted algorithms analyze dissimilarity and dynamic alterations of RAPD polymorphisms in the integrated micro-ecosystem *Cordyceps sinensis*. *Beijing Da Xue Xue Bao*. 2014;46(4):618–628.

42. Liang HH, Cheng Z, Yang XL, et al. Genetic variation and affinity of *Cordyceps sinensis* in Qinghai Province based on analysis of morphologic characters and inter-simple sequence repeat markers. *J Microbiol*. 2005;46(5):549–556.

43. Zhu ZX, Xie F, Zhang N. Genetic diversity of *Ophiocordyceps sinensis* in Gansu Province based on inter-simple sequence repeat (ISSR) analyses. *Mycosystema*. 2011;30(3):501–507.

44. Xiao YY, Cao YP, Zhang L, et al. Study on genetic diversity of anamorphic isolates from different sources of *Ophiocordyceps sinensis* by ISSR. *J Fungal Res*. 2014;12(3):154–159.

45. Sharon I, Banfield JF. Microbiology. Genomes from metagenomics. *Science*. 2013;342(6162):1057–1058.

46. Ye DD, Fan MM, Guan Q, et al. A review on the bioinformatics pipelines for metagenomic research. *Dongwuxue Yanjiu*. 2013;33(6):574–585.

47. Behzad H, Gojobori T, Mineta K. Challenges and opportunities of airborne metagenomics. *Genome Biol Evol*. 2015;7(5):1216–1226.

48. Jiang Y, Yao YJ. A review for the debating studies on the anamorph of *Cordyceps sinensis*. *Mycosistema*. 2003;22(1):161–176.

49. Nei M, Li WH. Mathematical model for studying genetical variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A*. 1979;76(10):5269–5273.

50. Odum EP. *Principles and concepts pertaining to organization at the community level*. In: Odum EP, editor. Philadelphia, USA: Fundamentals of ecology. W.B. Saunders; 1971. p. 140–161.

51. Guo Y, Xiao P, Wei JC. On the biology and sustainable utilization of the Chinese medicine treasure *Ophiocordyceps sinensis* (Berk.) Sung, G.H. et al. *Modern Chin Med*. 2010;12(11):3–8.

52. Dong YZ, Zhang LJ, Wu ZM, et al. Altered proteomic polymorphism in the caterpillar body and stroma of natural *Cordyceps sinensis* during maturation. *PLoS One*. 2014;9(10):e109083.